-/ CS170-1

Medical Care

January 1982
Vol. 20 No. 1
MDLCBD · ISSN 0025-7079

OFFICIAL JOURNAL OF THE MEDICAL CARE SECTION,
AMERICAN PUBLIC HEALTH ASSOCIATION

The Efficacy of Second-Opinion Consultation Programs: A Cost-Benefit Perspective

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Diseases

Assessing Medical Practices Through PSRO Cooperative Studies

Periodic Medical Review: The Impact in Pennsylvania

Alcoholism, Morbidity and Care-Seeking

PUBLISHED BY

J. B. Lippincott Company · Philadelphia · Toronto

EDITORIAL BOARD

Donald C. Riedel, Ph.D.
Editor-in-Chief
University of Washington

Jan Van Dusen
Editorial Assistant

S. E. BERKI, M.A.
University of Michigan

ROBERT H. BROOK, M.D., Sc.D.
University of California
Los Angeles

GORDON H. DE FRIESE, PH.D.
University of North Carolina

EUGENE FEINGOLD, PH.D.
University of Michigan

BRANT E. FRIES, PH.D.
Yale University

THOMAS R. GODKINS, B.S., P.A.
University of Oklahoma

HAROLD HUNTER, DR. P.H.
Washington, D.C.

SHERMAN A. JAMES, PH.D.
University of North Carolina

ANTHONY L. KOMAROFF, M.D.
Brigham and Women's
Hospital, Boston

JAMES P. LOGERFO, M.D., M.P.H.
University of Washington

NANCY MILIO, PH.D.
University of North Carolina

DUNCAN NEUHAUSER, PH.D.
Case Western Reserve University

JOHN E. WARE, JR., PH.D.
The Rand Corporation

ALONZO S. YERBY, M.D.
Harvard University

ALFRED YANKAUER, M.D.
University of Massachusetts
Medical School

MEDICAL CARE (ISSN 0025-7079) is published monthly by J. B. Lippincott Company, East Washington Square, Philadelphia, PA 19105. Subscription address: Journal Fulfillment Department, Lippincott/Harper, 2350 Virginia Avenue, Hagerstown, MD 21740. Subscription rates: United States and U.S. Possessions, one year, $40.00; special price U.S. students, $23.00; all other countries except Japan, one year, $50.00. Single copies $5.00. Subscription orders and correspondence concerning subscriptions in Japan should be sent to Woodbell Scope, 11-11 Shoto 2-chome, Shibuya-ku, Tokyo 150, Japan. Printed in U.S.A. Copyright © 1982 by J. B. Lippincott Company. Second class postage paid at Hagerstown, Maryland, and additional mailing offices.

January 1982
Volume 20, Number 1

CONTENTS

EDITORIAL

Second-Opinion Programs: Beyond Cost-Benefit Analyses. *Robert H. Brook and Kathleen N. Lohr* 1

ORIGINAL ARTICLES

The Efficacy of Second-Opinion Consultation Programs: A Cost-Benefit Perspective. *Hirsch S. Ruchlin, Madelon Lubin Finkel and Eugene G. McCarthy* 3

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates. *Suzanne Grisez Martin, Michael Shwartz, Bernadette J. Whalen, Deborah D'Arpa, Greta M. Ljung, John Holden Thorne and Anne E. McKusick* 21

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization. *Joseph D. Restuccia* 46

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Disease. *James L. Rogers, Olga M. Haring, Paul M. Wortman, Royce A. Watson and John P. Goetz* 63

Assessing Medical Practices Through PSRO Cooperative Studies. An Evaluation of Cesarean Births in Nine PSRO Areas. *William F. Jessee, C. W. Nickerson and William S. Grant, for the PSRO Coordinating Group* 75

Periodic Medical Review: The Impact in Pennsylvania. *Michael A. Calogero and Daniel B. Hill* 85

Alcoholism, Morbidity and Care-Seeking. The Inpatient and Ambulatory Service Utilization and Associated Illness Experience of Alcoholics and Matched Controls in a Health Maintenance Organization. *Sandra L. Putnam* 97

LETTER TO THE EDITOR

On Assessing the Performance of Medical Care Systems. *James I. Hudson.* 122

BOOK REVIEWS

Toward Rational Technology in Medicine: Considerations for Health Policy, by H. David Banta, Clyde J. Behney and Jane Sisk Willems. *Eric B. Larson* 122

Assessing the Contributions of the Social Sciences to Health, edited by M. Harvey Brenner, Anne Mooney and Thomas J. Nagy. *Robert L. Eichhorn* 123

The Planning of Health Services: Studies in Eight European Countries, edited by G. McLachlan. *L. S. Rosenfeld* 124

Women, Work, and Health: Challenges to Corporate Policy, edited by Diana Chapman Walsh and Richard H. Egdahl. *Gail Gordon* 125

MEDICAL CARE
January 1982, Vol. XX, No. 1

Original Articles

The Efficacy of Second-Opinion Consultation Programs: A Cost–Benefit Perspective

HIRSCH S. RUCHLIN, PH.D.,* MADELON LUBIN FINKEL, PH.D.,† AND EUGENE G. MCCARTHY, M.D., M.P.H.‡

This study evaluates a mandatory second-opinion consultation program administered on behalf of a large Taft-Hartley welfare fund providing medical care coverage for 120,000 beneficiaries and covered dependents. During a two-year intake period (1977–1978), 2,284 individuals received second-opinion consultations for an elective surgical procedure recommended by a first-contact physician or surgeon. Of this group, 366 received a nonconfirmation of their need for surgery. Medical claims data were available for 342 individuals in this group, and they constitute the base for the current analysis. A comparable number of individuals who received a positive confirmation were randomly selected and served as a control for estimating program savings. Both groups were followed for a one-year period from the date of their consultations. Total program savings were estimated at $534,791. Of this amount, medical care utilization savings were $361,756 and productivity savings were $173,035. The cost of the program was $203,300, yielding a benefit–cost ratio of 2.63. These findings indicate that mandatory second-opinion consultation programs, which are consumer oriented and intervene before care is rendered, are clearly cost-effective.

SECOND-OPINION CONSULTATION PROGRAMS for elective surgery are designed primarily to assist patients in making a more informed decision once elective surgery has been recommended. The principal goal of such programs is to provide patients with additional information on their need for surgery and, where applicable, on alternate modes of treatment. In addition to improving quality of care, second-opinion consultation programs, by reducing the rate of elective surgery, can also lead to significant economic savings.

Studies reported in the literature indicate that approximately 30 per cent of the individuals who voluntarily seek a second-opinion consultation, and approximately 18 per cent of those who are re-

---

This research was supported by Contract No. 600-75-0175 from the Health Care Financing Administration, Department of Health and Human Services.

* Professor of Economics in Public Health, Department of Public Health, Cornell University Medical College, New York, New York.

† Assistant Professor, Department of Public Health, Cornell University Medical College, New York, New York.

‡ Clinical Professor, Department of Public Health, Cornell University Medical College, New York, New York.

Address for reprints: Hirsch S. Ruchlin, Ph.D., Department of Public Health, Cornell University Medical College, 1300 York Avenue, New York, NY 10021.

0025-7079/82/0100/0003/$01.40 © J. B. Lippincott Co.

quired by their insurance carriers to seek a second opinion, receive a nonconfirmation§ of their need for surgery.[1–7] The magnitude of the nonconfirmation rate is taken as prima facie evidence of the presence of "surplus surgery." Follow-up studies of individuals participating in the largest mandatory second-opinion consultation program—the one conducted by Cornell University Medical College in conjunction with several Taft-Hartley security welfare funds‖ in the greater New York City area—indicate that of those not confirmed for surgery, 61 per cent had not had the surgery performed one year from the date of their second-opinion consultation. Moreover, 51 per cent of those not confirmed for surgery reported that they neither had the surgery performed nor received any form of medical treatment since their second-opinion consultation. For those who voluntarily sought a second-opinion consultation, these statistics are 78 per cent and 41 per cent, respectively.[8] On the basis of these data, one can hypothesize that second-opinion programs should yield significant economic savings. The current study presents data supporting this hypothesis.

In evaluating a second-opinion consultation program, one must recognize the fundamental limitations inherent in analyzing medical or economic outcome when the election of a second-opinion consultation is voluntary. Experience indicates that less than 5 per cent of those eligible for participation in a voluntary second-opinion consultation program take advantage of this benefit. By the very nature of the program's design, individuals who voluntarily seek a second opinion are a self-select group. Evaluations based on the results of a mandatory program do not suffer from such inherent selection biases and consequently have greater validity and generalizability.

The current study is based on the experience of a large mandatory second-opinion consultation program administered by the Building Service Local 32B-J welfare fund. This fund provides health benefit coverage for 120,000 individuals (employees and covered dependents) living in the greater New York City metropolitan region. The methodology used in the current study is outlined in the next section. This is followed by a description of the study population and the data base, and analyses of program savings and costs. The findings emanating from these analyses are presented and discussed in the concluding sections.

Methodology

The primary focus of the economic evaluation is to ascertain the cost-containment potential of a mandatory second-opinion consultation program. Consequently, the following rule was adopted in the modeling of the program: Any recommendation emanating from the consultation process that could potentially lead to reduced health care expenditures is considered to be a nonconfirmation of the original recommendation for surgery, e.g., surgery performed on an ambulatory basis. This is the case even if the new recommendation results in a modified surgical regimen or if the initial recommendation for surgery is ultimately accepted after trying a recommended medical treatment.

As a result of receiving a second-opinion consultation, individuals are classified into one of two categories: confirmed or not-confirmed.# The evaluation focus is thus centered on ascertaining whether the *receipt* of a second-opinion consultation generates economic savings (benefits). These savings should materialize through the election of either less surgery, less expensive surgery or less expensive medical care.¶

---

§ A nonconfirmation results when the consultant surgeon feels that the initially recommended surgery is not the most appropriate therapeutic next step. The consultant may feel that medical treatment or an alternate surgical procedure is more appropriate, or that there is no pathological justification for surgery.

‖ These funds were created by the Taft-Hartley Act of 1948 and are the principal mechanisms by which private industry in the United States provides fringe benefit coverage to workers and their dependents. Management and labor are equally represented in each fund.

Within both the confirmed and the not-confirmed groups, one can delineate various subgroups based on the individual's actions following the receipt of the second-opinion consultation. Individuals who received a confirmation of their need for surgery could elect or reject surgery. Not-confirmed individuals can be placed in one of four categories: those who had no surgery, those who elected a modified surgical procedure, those who elected the originally recommended surgery after trying an intervening medical treatment and those who elected the originally recommended surgery without any intervening medical care. Although the cost–benefit evaluation reported here will be based solely on data pertaining to the entire not-confirmed and confirmed groups, medical care utilization data for each subgroup are of interest and are reported in the Appendix.

Resources saved by the avoidance of surgery or the election of a less expensive treatment constitute the economic measure of the program's success. These savings are ascertained by subtracting the total cost profiles generated by the not-confirmed group from the total cost generated by the confirmed individuals. If the program is successful, in an economic sense, the not-confirmed group will exhibit lower expenditure patterns than the confirmed group. Two major benefit categories were delineated for the current study: 1) savings of health-related resources and 2) avoidance of work loss and restricted-activity days. The first type of saving encompasses reduced surgical and medical care, hospitalization, drug and therapy use, and ancillary service use attributable to the second-opinion consultation program. Savings stemming from reduced work loss and restricted-activity days avoided reflect the indirect benefits gained. The cumulative monetary value of both types of benefits is compared with the cost of the second-opinion consultation program. Program cost consists of the administrative costs incurred to operate the program, the expenditures for the second-opinion consultation visit and any ancillary tests ordered as part of the consultation, travel costs incurred by program participants in obtaining the consultation and wage loss (i.e., opportunity costs) due to work time missed in order to obtain the second-opinion consultation.

Findings reported in this study identify utilizer and cohort patterns. Data presented for utilizers indicate only specifically *reported* services consumed and work loss and restricted-activity days incurred. Cohort data represent actual reported utilization and, where necessary, proxied values. Means reported for the cohorts were derived by summing all reported and proxied data and dividing this statistic by the total number of utilizers and *nonutilizers* in that group. These means indicate average expected utilization and cost spread over the entire group. Such a

---

Individuals participating in the second-opinion program may request an additional (i.e., third) consultation. In such cases, the opinion of the last (i.e., second) consultant determines whether the intial recommendation for surgery is considered to be confirmed.

¶ One should not expect every "not-confirmed" individual to decide not to have surgery. Similarly, not every "confirmed" individual will elect to have surgery. However, one expects that the majority of individuals in each group will follow the consultant's recommendation. For the current study population, 89 per cent of the confirmed individuals elected surgery, while 85.1 per cent of the not-confirmed group followed the recommendation of the second-opinion consultant. As the program being evaluated permits the individual to decide whether or not to elect surgery, both behavioral patterns must be recognized within the confirmed and not-confirmed groups.

statistic has relevance for actuarial and insurance analyses. Means reported for the utilizers depict average utilization and cost incurred only by those individuals actually reporting data. As both the actuarial and actual utilization perspectives are of interest, they are both reported.

In analyzing the economic merit of a second-opinion consultation program, one should focus on total group (i.e., cohort) utilization and cost patterns. Within any group, savings can emerge both from changes in medical care utilization patterns among those who utilized care and from the avoidance of care. The latter phenomenon can only be seen if total group (i.e., cohort) patterns are analyzed. The same rationale dictates that groupwide patterns of work loss and restricted-activity days are pertinent for the current study. Consequently, the cost–benefit analysis focuses primarily on cohort-specific data.

Study Population

The study sample was drawn from the population of all covered individuals who were advised to have elective surgery during the period from January 1, 1977 to December 31, 1978 (n = 2,284). Of this group, 366 individuals did not have their need for surgery confirmed by a consultant surgeon. A comparable number of individuals whose need for surgery was confirmed by a consultant surgeon were randomly selected from the total number of confirmed individuals (n = 1,918).** Data could not be obtained for 24 not-confirmed individuals; thus, the final study population consisted of 342 not-confirmed and 342 confirmed individuals.††

The median age category for both the confirmed and not-confirmed groups is 45–64 years. Approximately 7 per cent of the total study population were in age category 15–24 years, 37 per cent were in age category 25–44 years, 53 per cent were in age category 45–64 years and 3 per cent were 65 years of age or older. Fifty-two per cent of the study population were male, and 48 per cent were female. Seventy-five per cent of the individuals included in the study were married, 13 per cent were single and 12 per cent were divorced or separated. Approximately 67 per cent were employed within the six month period following the second-opinion consultation, and an additional 25 per cent were homemakers (housewives). Only 1 per cent of the study population were unemployed. The remaining study participants (6 per cent) were students, retirees or permanently disabled. Approximately half of the individuals in both the confirmed and not-confirmed populations reported having completed less than 12 years of formal schooling, and an additional 14 per cent indicated that they did not know the extent of their educational attainment. Only 2 per cent had completed college. Two thirds of the employed respondents in both the confirmed and not-confirmed groups reported weekly earnings in the $200–299 range.

Ten surgical procedures account for approximately 60 per cent of the initial recommended surgeries: eye surgery except cataract (9.3 per cent); mass, cyst or lesion excision other than varicose veins (8.8 per cent); hernia repair (8.8 per cent); hysterectomy (7.4 per cent); dilation and curettage (5.4 per cent); incision, excision and suture for mass, cyst or lesion of the skin (4. 8 per cent); repair of deviated septum (3.9 per cent); foot surgery except bunionectomy (3.8 per cent); removal of cataract (3.6 per cent); and mastotomy, mastectomy and other operations of the breast (3.5 per cent).

Each person included in the cost–benefit analysis was followed for a period of one year following their second-opinion

---

** The data pool available for analysis did not permit a case-by-case match by diagnosis.

†† This sample size is greater than that required at $p \leq 0.05$ to note signifiant program savings.

consultation. As entry into the study spanned a two-year period (January 1, 1977 to December 31, 1978), cost data spanned a three-year period—1977 through 1979. All 1978 and 1979 cost data were deflated to a 1977 base, utilizing a 10 per cent discount rate.

Data Base

Data Sources and Proxies

Data utilized in this study were drawn primarily from welfare fund medical claim records and secondarily from telephone-administered questionnaries administered to all study participants one year after the anniversary date of the second-opinion consultation. Interviews were utilized to ascertain the amount of work loss and restricted-activity days, and out-of-pocket medical care costs. Data on medical care utilization were also gathered through the questionnaire and used in the analysis in instances where the person's claim file was incomplete. Data on the cost of the program were obtained from records maintained by the fund.

In generating the data base for this study, extreme care was exercised to ensure that medical utilization, work-loss and restricted-activity data related *only* to the condition for which surgery was originally recommended were included. Medical utilization, employment and restricted-activity data resulting from other causes were *excluded* from the study.

Ninety-two per cent of the study population (629 people) responded to the cost–benefit questionnaire. As medical claim form data were available for nonrespondents, it was decided to include them in the study and to proxy the missing data. Where precise information was not reported for the cost of office-based physician visits, the price of such visits was proxied at $30/visit. This proxy value was selected as it represented the prevailing office visit charge in the New York City area during the time of the study. In all cases, hospital costs were estimated based on the prevailing Blue Cross and Blue Shield of Greater New York per diem reimbursement. Hospital charge data were not used since charges overestimate the true cost of rendering hospital care. As the Blue Cross policy offered by the welfare fund provides full coverage for the first 21 days of a hospital stay and half coverage for the succeeding 180 days, the Blue Cross payment for care beyond the 21st day was doubled. Charges for medically related services not covered by Blue Cross, such as the cost of blood, were added to the reimbursement revenue to obtain the aggregate cost of hospital care.

In cases in which a specific drug cost was not provided, but in which the respondent reported the type of drug prescribed and the frequency and duration of intake, drug costs were proxied using prices listed in the *Lilly Drug Directory*. In the case of eye surgery, the cost of eye drops was proxied at $1 per month, the cost of eye glasses at $75/pair and the cost of contact lenses at $250/pair. These estimates were based on expert medical opinion. If direct proxies could not be obtained, mean values for the particular data element based on information provided by the applicable population group (i.e., confirmed and not-confirmed) was used as an indirect proxy.

Data on individuals' earnings were required to calculate the value of incurred work losses. Where necessary, wage data for Building Service union members were proxied based on collective bargaining hourly wages for the pertinent occupational category. For non-Building Service spouses, proxies were derived, where necessary, from Bureau of Labor Statistics regional wage reports.

The price used to value homemaker (housewife) time was based on the market valuation of homemaker services generated by Wendyce Brody.[9] Brody estimated the annual value of housewife services in

1972 as $4,705. This estimate was based on market wages prevailing in the Syracuse, New York area. Bureau of Labor Statistics wage level indices were used to: a) adjust the data to reflect wage patterns prevailing in the New York City region and b) inflate 1972 data to reflect market values prevailing during the current study period.

The prevailing minimum wage was used to value the restricted-activity time of students, retirees, permanently disabled, and laid-off or unemployed individuals. If an individual's employment status changed during the study period (i.e., person retired, was laid off, etc.), the value of his/her lost time (i.e., actual or minimum wage) was determined by the employment status at the time the loss occurred.

Biases Inherent in the Data Base

Factors unique to the welfare fund study population and to the New York City health care system introduce a downward bias in some of the cost data utilized in this study. Specifically, surgeon and physician visit costs may be biased in a downward direction due to three factors: 1) the welfare fund's surgical fee negotiation program, 2) the availability of "panel" surgeons, and 3) the billing practices of the welfare fund's health center. The surgical fee negotiation program attempts to obtain reductions in surgical fees in cases where the surgeon accepts assignment and his fee exceeds the welfare fund's stipulated reimburement for any procedure.[10] Under this program, surgeons are initially requested to accept the welfare fund's reimbursement as payment in full or, if this is not acceptable, to reduce their fee voluntarily. In 7.7 per cent of the surgical cases included in the cost–benefit study, the surgeon's fee initially exceeded the welfare fund's allowable scale, and negotiation occurred.

Some individuals electing surgery may have used a "panel" surgeon. If a beneficiary, having gone for a second-opinion consultation, does not have a surgeon to perform the elective surgery, the welfare fund will arrange for a "panel" surgeon to perform the surgery. Panel surgeons are board certified and agree to accept the welfare fund's designated reimbursement. Furthermore, panel members are not permitted to bill the welfare fund or the beneficiary for routine postoperative medical care (5.7 per cent of the people electing surgery had their surgery performed by a panel surgeon).

Building Service Local 32B-J also operates a health center for its members. Fees at the health center are lower than those prevailing in the city at large (e.g., the cost of a physician visit is only $25). Furthermore, if a patient receives medical care there, the welfare fund will pay only 80 per cent of the medical bill after the annual $100 deductible is met. Health center physicians quite often fail to bill the patient for the remaining 20 per cent.

New York City has the largest network of municipal hospitals in the nation. Although municipal hospitals are frequently thought of as places that treat only the poor, they do render a substantial amount of care to the working population of the city. Two biases are introduced if care was obtained in the municipal system. Cost of outpatient department care may be biased downward as charges for such care are based on the patient's ability to pay. Furthermore, when a patient is hospitalized in a municipal hospital, he/she is treated by a salaried surgeon who is not permitted to bill the patient for this care. The cost of these services is presumably part of the total hospital bill. The same applies to the cost of anesthesiology.

A comparable situation exists when a patient is hospitalized in one of the city's major teaching hospitals. If the patient is placed on the teaching, as opposed to the private, service, a separate surgeon's charge is not forthcoming because the cost of medical and surgical care is presumably included in the aggregate hospital bill. Finally, in a number of cases, anesthesiologists did not bill the patient directly but included their fees as part of the surgeon's bill. In such cases (n = 61), the cost of surgical care is overestimated and the cost of anesthesiology care is underestimated.

The net effect of all these biases probably leads to an *underestimate* of the true savings of the second-opinion consultation program. The major impact of these biases is to reduce the cost of surgery, thereby underestimating program benefit. One should recognize that comparable biases would result in any study focused either on an urban lower-class working population or on special population groups such as Medicare and Medicaid patients, whose payment for care may be below prevailing physician charges.

Program Savings

Nine medical resource cost categories were delineated for this study: physician visits; hospital utilization; surgical fees for the original surgery; anesthesia fees for the original surgery; surgical fees for related surgery‡‡; anesthesia fees for a related surgery; diagnostics and laboratory; therapies and nursing; and drugs and prostheses. In addition, data are available on the number of physician visits, number of hospitalizations, total number of days of hospital care rendered and number of surgeries performed.

Medical resource utilization patterns displayed by the not-confirmed and confirmed groups are reported in Table 1. On

TABLE 1. Medical Resource Utilization Profiles

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Utilizers		Cohort			Utilizers	
Category	Total	Mean	Standard Deviation	Number	Mean	Total	Mean	Standard Deviation	Number	Mean
Number of physician visits	1,206	3.5	7.1	185	6.5	1,077	3.1	6.6	175	6.0
Number of hospital stays	325	0.9	0.6	298	1.1	151	0.4	0.6	133	1.1
Length of stay—all hospitalizations	2,130	6.2	7.7	298	7.2	1,206	3.5†	7.0	133	9.0
Number of surgeries*	396	1.2	0.8	304	1.3	193	0.6†	0.9	132	1.5

* This category includes the surgical procedures for which the individual received a second-opinion consultation (called original surgery) and all other surgical episodes related to the medical condition for which a second-opinion consultation was sought.

† Statistically significant from the mean value reported for the confirmed group at $p \leq 0.01$.

---

‡‡ In most cases, related surgery is a surgical procedure associated with the diagnosis for which a second opinion was sought. It is not, however, the procedure that was initially recommended. For example, the procedure initially recommended was dilation and curettage. This procedure was performed, and subsequently a hysterectomy was required. In other cases, related surgery refers to a diagnostic procedure, such as scoping, that is performed in a hospital under anesthesia and is done as a prelude to surgery, e.g., cystoscopy prior to a transurethral resection (prostatectomy).

both a cohort and utilizer basis, fairly comparable patterns emerge with regard to number of physician visits. Individuals reporting the use of physician care had approximately six visits in the follow-up year. On a cohort basis, utilization falls to approximately three visits per person per year. The number of hospital stays for those individuals reporting hospitalization was the same—1.1 stays per person. While hospital utilization among the not-confirmed group appears to be much lower on a cohort basis (0.4 versus 0.9 hospital stays), this difference fails to attain statistical significance at $p \leq 0.05$. Statistically significant differences do emerge with regard to cohort profiles for length of hospital stay and number of surgeries. For both of these phenomena, average use by the entire confirmed group was almost double that of the entire not-confirmed group—6.2 versus 3.5 hospital days, and 1.2 versus 0.6 surgeries.

Cost profiles generated from data reported by confirmed and not-confirmed individuals indicate that, on average, individuals in the confirmed group spent more money on medical care than their counterparts in the not-confirmed group, as can be seen from the data displayed in Table 2. Average cost per utilizer in the confirmed group was $2,429, as compared with $1,850 for the not-confirmed group. For those individuals utilizing medical care, mean cost of physician care, anesthesiology (related surgery), diagnostics and laboratory, and therapies and nursing was higher for the confirmed than for the not-confirmed group. For the remaining five categories—hospital care, surgical fees (original surgery), anesthesiology fees (original surgery), surgical fees (related surgery), and drugs and prostheses—the reverse was true.

Cohort-wide cost patterns also indicate that average cost per person in the not-confirmed group was less than that reported for the confirmed group. Total medical care expenditures incurred by the entire not-confirmed group equaled $425,538, contrasted with $787,294 for the entire confirmed group. Viewed from another perspective, average medical care cost per cohort member was $1,244 for the not-confirmed group and $2,302 for the confirmed group. As expected, two categories account for the majority of medical care expenditures for both cohorts. These categories are hospital utilization and surgical fees (original surgery). Jointly they account for 84 per cent of the expenditures incurred by the confirmed group and 79 per cent of the expenditures incurred by the not-confirmed group. On a per-cohort-member basis, differences in mean expenditures for hospital utilization were $635 ($1,348 – $713), and differences in mean expenditures for surgical fees for the original surgery were $324 ($601 – $277). It is interesting to note that statistically significant differences in total group expenditures appear for two additional medical resource categories—anesthesiology fees (original surgery) and therapies and nursing. In both cases, expenditures incurred by the confirmed group exceeded those incurred by the not-confirmed group.

Seven productivity-loss categories were delineated for this study: work/home/school loss due to days spent in the hospital (labeled "hospitalization"); posthospitalization regular work (labeled "main job") loss; overtime work loss§§; work loss at any additional jobs; posthospitalization restricted-activity days incurred by students, retired, disabled or unemployed individuals; and work-loss days incurred by relatives (other than spouses) of patients who had to attend to the needs of the patient. In addition, data were gathered on expenditures for outside household help that had to be hired to substitute for an ill housewife, and time loss (i.e., travel, wait-

---

§§ If a spouse incurred work-loss days to attend to the needs of the patient, this loss was included in the main job and overtime—main job categories.

TABLE 2. Medical Resources Cost Profiles

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Utilizers		Cohort			Utilizers	
Category	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)
Physician visits	31,074	91	207	185	166	27,902	82	185	175	156
Hospitalization	460,979	1,348	2,013	298	1,545	243,819	713*	1,604	133	1,821
Surgical fees—original surgery	200,068	601†	496	296	676	92,844	277*‡	479	120	774
Anesthesiology fees—original surgery	50,342	185§	131	217	214	20,692	68*‖	134	86	230
Surgical fees—related surgery	15,661	46	217	24	652	21,090	62	329	23	901
Anesthesiology fees—related surgery	4,343	13	61	19	229	4,827	14	62	20	221
Diagnostics and laboratory	10,446	31	166	49	212	5,361	16	54	50	106
Therapies and nursing	7,447	22	162	13	570	1,368	4#	38	6	223
Drugs and prostheses	6,934	20	92	78	85	7,635	22	173	65	115
Total cost¶	787,294	2,302	2,741	322	2,429	425,538	1,244*	2,427	228	1,850

* Statistically significant from the mean value reported for the confirmed group at $p \leq 0.01$.

† The number of observations used to calculate this mean was 333. Nine individuals in this group did not have separate, identifiable surgical bills.

‡ The number of observations used to calculate this mean was 334. Eight individuals in this group did not have separate, identifiable surgical bills.

§ The number of observations used to calculate this mean was 272. Seventy individuals in this group either did not incur any anesthesiology costs or did not have separate bills for anesthesiology.

‖ The number of observations used to calculate this mean was 306. Thirty-six individuals in this group either did not incur any anesthesiology costs or did not have separate bills for anesthesiology.

Statistically significant from the mean value reported for the confirmed group at $p \leq 0.05$.

¶ The number of utilizers in the total cost row exceeds the number in any previous row. This is due to the fact that any individual may report data for more than one category, but not for all categories. Thus the statistic listed in the total cost row indicates the number of individuals who reported *any* expense.

TABLE 3. Productivity Losses (in days)

Category	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Reported		Cohort			Reported	
	Total	Mean	Standard Deviation	Number	Mean	Total	Mean	Standard Deviation	Number	Mean
Hospitalization	2,130	6.2	7.7	296	7.2	1,209	3.5*	7.0	131	9.2
Main job	4,940	14.5	21.8	158	28.3	2,115	6.2*	20.1	85	23.4
Overtime (in hours)	556	1.6	8.4	19	27.1	401	1.2	7.7	13	25.4
Additional job	28	0.1	1.2	2	13.0	0	—	--	0	—
Homemaker	1,846	5.4	21.9	46	39.4	902	2.6	17.0	24	35.9
Restricted activity	1,097	3.2	22.3	18	55.9	1,267	3.7	26.9	12	99.8
Relative work loss	400	1.2	12.3	24	15.5	83	0.2	1.4	11	6.6

* Statistically significant from the mean value of the confirmed group at $p \leq 0.01$.

ing and examination time) and travel costs incurred for obtaining medical care in the year following the second-opinion consultation.

Productivity-loss patterns are presented in Table 3. On average, individuals in the confirmed group who reported actual productivity losses sustained 7.2 work/home/school loss days due to hospitalization and 28.3 main job work-loss days in the posthospitalization period. Not-confirmed individuals sustained 9.2 work/home/school loss days due to hospitalization and 23.4 posthospitalization main job work-loss days. Housewives (homemakers) in the confirmed group reported an average home loss of 39.4 days, while those in the not-confirmed group reported home losses averaging 35.9 days.

When viewed from the perspective of the entire cohort, much greater mean productivity losses emerge. The entire not-confirmed group averaged approximately half as many work/home/school loss days due to hospitalization and posthospitalization work-loss days as the entire confirmed group. Averaged over the entire cohort, confirmed group members lost 6.2 work/home/school days due to hospitalization, while not-confirmed group members lost only 3.5 days. Furthermore, confirmed individuals averaged 14.5 posthospitalization work-loss days, while not-confirmed individuals lost only 6.2 work days. Both the hospitalization and posthospitalization work-loss patterns are statistically significant at $p \leq 0.01$. Differences with regard to the remaining productivity-loss categories (i.e., overtime loss, additional job loss, homemaker loss days, restricted-activity days and relative work-loss days) were not statistically significant at $p \leq 0.05$.

The value of the total productivity loss sustained by the confirmed and not-confirmed groups is reported in Table 4. The entire not-confirmed group productivity loss was $175,621, contrasted with $348,656 for the entire confirmed group, resulting in a statistically significant program saving of $173,035. The three major productivity cost components, accounting for over 80 per cent of the total cost incurred by each group, were posthospitalization main job work loss, hospitalization work/home/school loss, and posthospitalization home-loss days incurred by housewives. In each of these cases, average group losses ($539 versus $243, $202 versus $105 and $144 versus $69) were significantly different at p ≤ 0.05 or lower.

By combining medical resource expenditures and productivity-loss patterns, one obtains the total economic costs incurred by the not-confirmed and the confirmed groups. Total economic costs incurred by the not-confirmed cohort in the one-year follow-up period were $601,159 ($425,538 in medical care costs and $175,621 in productivity losses), as compared with $1,135,950 ($787,694 for medical care and $348,656 in productivity losses) incurred by the confirmed group. Subtracting the costs incurred by the not-confirmed cohort from those incurred by the confirmed cohort yields a total program saving (i.e., benefit) of $534,791. Sixty-eight per cent of this total saving resulted from decreased use of medical resources and 32 per cent from decreased productivity losses.

Program Cost

Program costs consist of the administrative costs incurred to operate the program, the expenditures for the second-opinion consultation visits and any ancillary tests ordered as part of the consultation, and out-of-pocket (i.e., travel) and opportunity (i.e., wage loss) costs incurred in the process of obtaining the consultation.

The costs of administering the second-opinion program during 1977 were estimated from records maintained by the Building Service Local 32B-J welfare fund. Personnel costs totaled $29,350, of which $19,350 were salary and fringe benefit

TABLE 4. Cost of Productivity Loss

	Confirmed (N = 342)					Not-Confirmed (N = 342)				
	Cohort			Reported		Cohort			Reported	
Category	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)	Total ($)	Mean ($)	Standard Deviation	Number	Mean ($)
Hospitalization	69,030	202	249	293	236	36,063	105*	232	129	278
Main job	184,193	539	814	158	1,056	82,953	243*	934	85	899
Overtime	4,575	13	76	19	224	2,846	8	56	12	221
Additional job	1,121	3	55	2	530	0	—	—	0	—
Homemaker	49,173	144	581	46	1,049	23,751	69†	453	24	954
Restricted activity	16,566	48	361	18	906	20,023	59	435	12	1,616
Relative work loss	13,333	39	283	20	437	2,352	7†	44	11	185
Outside household help	882	3	38	2	437	735	2	28	2	367
Time lost for obtaining medical care	6,735	20	25	252	22	4,681	14*	23	171	22
Cost of travel for obtaining medical care	3,048	9	17	241	11	2,217	6	16	159	11
Total cost	348,656	1,019	1,385	315	1,030	175,621	514*	1,170	207	802

* Statistically significant from the mean value of the confirmed group at p ≤ 0.01.

† Statistically significant from the mean value of the confirmed group at p ≤ 0.05.

costs for 1.5 (FTE) intake officers, and $10,000 were allocated as the stipend of the welfare fund's medical advisor. Nonsalary costs totaled $10,000, of which $7,200 are imputed rental costs, and $2,800 represent actual incurred costs for office expenses. A separate study was not conducted to ascertain 1978 costs, which would have had to be deflated to a constant (1977) dollar level. As no change occurred in the administrative structure of the program during 1978, 1977 costs were used as a proxy for 1978 deflated costs. As program intake spanned a two-year period (January 1, 1977 through December 31, 1978), it is estimated that total administrative costs were $78,700.

During the two-year study period, 2,284 individuals obtained second-opinion consultations and 24 individuals obtained an additional (i.e., third) opinion. On average, the costs of the second-opinion consultation and associated ancillary procedures was $43.95, and the average cost of third-opinion consultations and associated ancillary procedures was $47.42. [Both of these cost statistics are in constant (1977) dollars.] Based on the number of consultations rendered, total expenditures for all consultations were $101,520.

Detailed data were not gathered on travel costs and opportunity costs incurred in obtaining second-opinion consultations. It is assumed that travel to and from the consultant's office as well as the examination took, on average, two hours. The average hourly wage prevailing for the study population, $5, was used to derive an aggregate travel and opportunity cost estimate of $23,080.

Combining the three program-cost elements yields a total program-cost estimate of $203,300. Fifty per cent of the costs were incurred for medical expenses associated with the second-opinion consultation, 39 per cent of the costs were accounted for by administrative expenses, and 11 per cent of the costs reflect out-of-pocket and wage loss incurred to obtain the consultation.

Cost–Benefit Calculations

Based on the experience of the not-confirmed and confirmed groups, which include both individuals who complied with the recommendation of the second-opinion consultant and those who did not, one obtains a benefit–cost ratio of 2.63. The components of this ratio are displayed in Table 5. This benefit–cost ratio implies that for every $1.00 of program cost incurred, $2.63 of benefits accrue.[11] An alternate evaluation statistic is the program's net present value (NPV). This is obtained by subtracting program cost from program benefit. The NPV for the second-opinion consultation program is $331,491. This indicates that net benefit (i.e., total benefit minus total cost) equaled $331,491 for the program under study. Dividing net benefit by the total number of program participants (2,284) yields a net benefit per participant of $145.14. Thus, on average, for every individual who is told to obtain a second-opinion consultation (a population pool much larger than the number not-confirmed), one can expect to save $145.14.

The perspective adopted in the current study recognizes benefits from both reduced utilization of medical resources and reductions in productivity loss. If only medical resource utilization that can lead to a direct reduction in medical care expenditures and cost borne directly by the program had been considered in the analysis, the ensuing benefit–cost ratio would have been 2.00. Even from this (narrow) perspective, the second-opinion consultation program must be considered

---

[11] In order to test the sensitivity of this finding to the selected discount rate, 10 per cent, the entire analysis was repeated using alternately rates of 15 per cent and 6 per cent. The resulting benefit–cost ratios were 2.53 and 2.72, respectively.

TABLE 5. Cost–Benefit Statistics

Component	Value
Total costs incurred by the confirmed group (B_e)	\$1,135,950
Total costs incurred by the not-confirmed group (C_e)	\$ 601,159
Program cost (U_e)	\$ 203,300
Number of participants (P)	2,284

Cost–Benefit Equations

$$\text{Benefit/Cost} = \frac{B_e - C_e}{U_e}$$

$$= \frac{1{,}135{,}950 - 601{,}159}{203{,}300}$$

$$= \frac{534{,}791}{203{,}300} = 2.63$$

$$\text{Net Present Value} = (B_e - C_e) - U_e$$

$$= 534,\ 791 - 203,\ 300 = \$331{,}491$$

$$\text{Net Present Value/Participant} = \frac{(B_e - C_e) - U_e}{P}$$

$$= \frac{331{,}491}{2{,}284} = \$145.14$$

cost-beneficial in that for every \$1.00 of cost incurred, \$2.00 of benefit (reduced medical resource utilization) occurred.##

Summary and Discussion

An economic evaluation of the mandatory second-opinion consultation program administered on behalf of the Building Service Local 32B-J welfare fund indicates that substantial net savings were reaped both by the welfare fund and its covered beneficiaries. Comparing the medical utilization histories of the not-confirmed and the confirmed groups during the first year after receipt of the consultation indicates that the not-confirmed group spent \$361,756 less on medical care than the confirmed group. Furthermore, due to the avoidance of surgery, the not-confirmed group exhibited fewer work-loss, home-loss and restricted-activity days. This saving, referred to as a productivity gain, was valued at \$173,035. Jointly, savings from reduced medical care utilization and increased productivity equalled \$534,791.

Total costs incurred by the second-opinion consultation program during the two-year study period (1977–1978) were \$203,300. Total expenditures for consultations and associated ancillary procedures

---

The computation of medical resource savings used throughout this study is based on a long-run perspective in which all costs are considered to be variable. In the short-run, some costs are fixed and will be incurred even when utilization declines. This is particularly true in the case of hospital care. As agreement does not exist as to the percentages of total costs that are fixed and variable, respectively, we assume that 50 per cent of costs are fixed and 50 per cent variable. Utilizing this assumption, short-run savings due to decreased hospital use decline from \$217,610 (i.e., \$460,979 – \$243,819; line 2, Table 2) to \$108,580, thereby reducing the aggregate benefit–cost ratio from 2.63 to 2.10, and the benefit–cost ratio based exclusively on reduced medical resource utilization declines from 2.00 to 1.25. It is our opinion, however, that a long-run perspective, in which all costs are variable, is the proper one for policy analysis.

equalled $101,520. Administrative costs were $78,700, and estimated out-of-pocket and opportunity costs were $23,080.

Merging the estimates of program benefit and cost yields a benefit–cost ratio of 2.63. That is, for every $1.00 of cost incurred, $2.63 of benefits occurred. Based exclusively on this "pay-back ratio," one must conclude that there are definite economic benefits to be derived from a second-opinion consultation program. While this evaluation did not focus on noneconomic benefits, the majority of individuals responding to questionnaires volunteered the information that they were feeling better during the postconsultation period.

Other evaluations of second-opinion consultation programs have suggested that the receipt of a second-opinion consultation is cost beneficial.[5,6,11] The results of the current study add further weight to this contention. Furthermore, the results of the current study should be viewed as more definitive, as the methodology and data base used in this study overcome many of the shortcomings inherent in previous evaluations of the economic potential of second-opinion consultation programs.

The first published evaluation of a second-opinion consultation program focused on the intial 20 months of Blue Cross and Blue Shield of Greater New York's voluntary second-surgical-opinion program. One of the major findings reported in this evaluation, based on the experience of over 1,500 program participants, was an estimated gross savings of $600,000 in health care costs. After deducting the costs of administering the program, which included relatively high start-up costs, the evaluators estimated a net savings due to the program of a quarter of a million dollars.[11] It should be recognized that this estimate excludes any productivity gains resulting from the avoidance of work-loss or home-loss days and is based exclusively on assumed gross savings from the avoidance of surgery. Thus, from a societal perspective, the reported savings are underestimates. In generating the assumed savings resulting from the avoidance of surgery, the relative experience of a "control" group similar to that specified in this study was not utilized. Therefore, savings reported from the avoidance of surgery must be viewed as gross savings that were not offset by expenditures incurred by individuals not electing surgery, and as such, these reported savings are overestimates of true medical resource savings.

A recently published evaluation of Massachusetts' mandatory consultation program for elective surgery for Medicaid recipients estimated that net statewide savings to Medicaid during the first year of program operation were $856,500, and when benefits are compared with costs, the program exhibits a benefit–cost ratio of 3.9.[5] It should be noted that this study considered program costs and savings strictly from Medicaid's perspective. If costs were borne by other parties, or if benefits accrued to other parties, they were not included in the analysis. Furthermore, data on medical care expenditures were based on estimated Medicaid payments rather than claim files data, which were used in the current study. Productivity gains or losses were omitted from the analysis of program benefit, and gross estimated savings from foregone surgery were offset only by the cost of other surgery, if other surgery was recommended by the consultant. Nonsurgical expenses, such as those for physician visits and drug therapy, were not estimated for individuals foregoing surgery and were not subtracted from estimated gross savings. As in the case of the Blue Cross and Blue Shield of Greater New York study, the reported program savings and cost–benefit ratio were biased by these omissions.

Gertman et al. present an alternate estimate of the economic value of the Massachusetts program, based exclusively on the experience of 49 patients living in the metropolitan Boston area who received second opinions reversing the recommendation for a hysterectomy. In this study, program benefit was also estimated solely by projecting savings from changes in the amount of surgery performed. Based on this procedure, a benefit–cost ratio of 2.27 was reported. Applying this methodology to the other seven procedures encompassed by the Massachusetts program,¶¶ Gertman et al. speculate that the benefit–cost ratio would be only 1.11.[6] the shortcomings inherent in the first study of the Massachusetts Medicaid program apply with equal force to the Gertman et al. study.

The economic evaluation presented in the current study does not suffer from the shortcomings inherent in these sudies. Our perspective recognizes all social costs, not just costs germane to one party such as a Medicaid program. The definition of program benefit used in the current study encompasses both medical resource and productivity savings. Furthermore, the medical resource savings reported in this study are based on actual data maintained in a welfare fund's medical claims files. Cost estimates were used only when proxy values were needed.

While the selection of a "control" group for the current study represents a distinct methodological improvement over previous evaluation efforts in this subject area, it is recognized that an optimal control group was not used. Data availability limitations and the widespread knowledge of the merit of "second opinions" precluded the selection of a different control group. Thus, a key question raised by critics of a second-opinion consultation program, namely, "Does the program recruit individuals to surgery?" cannot be answered by our analysis. An exhaustive search of the literature failed to turn up any data on the proportion of individuals who decide not to undergo elective surgery without the benefit of additional information provided by a second-opinion consultation. Discussions with various welfare fund administrators in the New York City area that are either self-insured, self-administer their health benefits program or have traditional third-party carrier relationships indicate that the surgical utilization experience of their membership and covered dependents did *not* increase after the adoption of a second-opinion consultation program. In light of this development we proceeded with the use of the control group selected for this study.

All economic evaluations of second-opinion consultation programs have been based on a limited follow-up period. Evidence gathered in the course of this and other evaluations indicates that approximately 90 per cent of all individuals who will elect surgery after a second-opinion consultation will do so within six months of the consultation. Thus, the vast majority of medical care costs and productivity losses will occur within a year following the receipt of the consultation. Furthermore, when program participants have been followed for longer periods (i.e., two years), no evidence has been uncovered to indicate that any new medical complications occur.

The cost–benefit analysis presented here indicates that the second opinion consultation program administered on behalf of the Building Service Local 32B-J welfare fund returned $2.63 in benefits for every $1.00 of cost incurred. Had the program required that adherence to the consultant's recommendation be a prerequisite for full medical coverage, as is the case with the second-opinion program offered by the Prudential Insurance Company of America; had it covered a labor force commanding higher wage levels; and had it been targeted to selective diagnoses, the benefit–cost ratio would have exceeded

---

¶¶ These procedures are tonsillectomy and adenoidectomy, cholecystectomy, hemorrhoidectomy, laminectomy or spinal fusion, meniscectomy, submucous resection or nasal septal repair, and excision of varicose veins.

2.63. Support for the first contention comes from one of the substudies (not reported here) conducted as part of our overall research effort. When the analysis was limited solely to individuals who followed the recommendation of the consultant (cohorts 1 and 7 in the Appendix table), the reported benefit–cost ratio was 3.18, indicating a saving of $3.18 for every $1.00 expended. As noted in the "Study Population" section of this article, the median weekly salary for Building Service employees was in the $200–$299 range, implying that median hourly earnings are in the $5–$8 range. Had the covered employees commanded hourly wages in the $12–$16 range, as in the case for carpenters and teamsters, productivity savings would have been higher. This is the case because work-loss days avoided would have been valued at the higher wage level. Greater productivity savings would automatically lead to a higher benefit–cost ratio. Finally, there is reason to believe that economic benefits could be maximized by confining second-opinion consultation programs to those diagnoses that have the highest nonconfirmation rates. While additional research is needed to identify these diagnoses, the data reported in the evaluation of the Massachusetts program indicate substantial differences in net savings for each of the eight procedures relevant to the Massachusetts Medicaid program.[5] [***]

Conclusion

Second-opinion consultation programs represent the first concerted cost-containment effort targeted directly at consumers and individual providers of medical care. Prior cost-containment efforts have focused almost exclusively on the institutional providers of care. While early efforts at controlling cost via rate regulation[12–14] and utilization review[15–19] were not overly successful, these efforts are becoming more successful as the respective programs have matured.[20,21] To date, a comprehensive evaluation of rate regulation from a cost–benefit perspective has not been undertaken. While it has been shown that rate regulation reduces the rate at which hospital costs increase in a number of large industrial states, one is not sure whether the gains are being offset by the cost of the rate setting activity. Furthermore, it should be evident that rate regulation will succeed only in those states that have the political will to enforce and sustain such a system.

It is reasonable to assert that the control of health care costs requires a multiplicity of efforts. Although control of institutional costs is beginning to yield positive results, one must also develop programs that are directed toward the individual consumers and their physicians, who play a dominant role in determining the volume and type of medical care services that are consumed. To date, the PSRO program, which by design entails a *retrospective* review, has offered few direct challenges to the individual physician and does not attempt to interact directly with consumers. Second-opinion consultation programs are currently the only major *prospective* effort to influence the consumer and, through the consumer, to influence physician behavior indirectly. As such programs can both reduce surplus surgery and effectively contain costs, they merit inclusion in an array of public or private sector cost-containment efforts.

References

1. McCarthy EG, Widmer GW. Effects of screening by consultants on recommended elective surgical procedures. N Engl J Med 1974;291:1331.

2. Grafe WR, et al. The elective surgery second opinion program. Ann Surg 1978;188:323.

[***] Reporting on savings estimated for patients whose need for surgery was reviewed by the Bay State Health Care Foundation, Martin et al. estimate an aggregate benefit–cost ratio of 4.2. When procedure-specific estimates were made, the benefit–cost ratio ranged from a low 0.9 for cholecystectomy to a high of 7.8 for excision of varicose veins.

3. McCarthy EG, Finkel ML. Second consultant opinion for elective surgery programs: outcome status over time. Med Care 1978;16:984.

4. McCarthy EG, Finkel ML. Second consultant opinion for elective gynecologic surgery. Obstet Gynecol 1980;56:403.

5. Martin SG, et al. The effect of a mandatory second opinion program on Medicaid surgery rates: an analysis of the Massachusetts consultation program for elective surgery. Boston: Department of Public Welfare, Commonwealth of Massachusetts, 1980. (Mimeographed).

6. Gertman PM, et al. Second opinions for elective surgery: the mandatory medicaid program in Massachusetts. N Engl J Med 1980;302:1169.

7. Joffe J, Schachter M. Program for elective surgical second opinion: surgical experience of program participants, 1976-77. (PRESSO Report No. 3). New York: Health Affairs Research Department, Blue Cross and Blue Shield of Greater New York, 1980. (Mimeographed).

8. McCarthy EG, Finkel ML, Ruchlin HS. Second opinion elective surgery. Boston: Auburn House, 1981.

9. Brody WH. Economic value of a housewife. Research and Statistics Note 9-1975. DHEW Publication No. (SSA) 75-11701, August 28, 1975.

10. Kamons AS, et al. Cost containment by a third party payer: negotiations of surgical fees. Employee Benefits Journal 1980;5:2.

11. Joffe J. Programs for elective surgical second opinion evaluation: preliminary findings, January 1976–August 1977. (PRESSO Report No. 1). New York: Health Affairs Research Department, Blue Cross and Blue Shield of Greater New York, 1978. (Mimeographed).

12. Berry RE. Prospective rate reimbursement and cost containment: formula reimbursement in New York. Inquiry 1976;13:288.

13. Hellinger FJ. Prospective reimbursement through budget review: New Jersey, Rhode Island, and Western Pennsylvania. Inquiry 1976;13:309.

14. Baker FD. Prospective rate setting in Washington State. Topics in Health Care Financing 1976;3:57.

15. Flashner BA, et al. Professional standards review organizations: analyses of their development and implementation based on a preliminary review of the hospital admission and surveillance program in Illinois. JAMA 1973;223:1473.

16. Brian E. Foundation for medical care control of hospital utilization: CHAP—A PSRO protype. N Engl J Med 1973;288:878.

17. Westphal M, Frazier E, Miller MC. Changes in average length of stay and average charges generated following institution of PSRO review. Health Serv Res 1979;14:253.

18. Lave JR, Leinhardt S. An evaluation of a hospital stay regulatory mechanism. Am J Public Health 1976;66:959.

19. Brook RH, Williams KN. Evaluation of the New Mexico peer review system, 1971 to 1973. Med Care 1976;14(Supplement):

20. Biles B, Schramm CJ, Atkinson JG. Hospital cost inflation under state rate-setting programs. N Engl J Med 1980;303:664.

21. Health Care Financing Administration, Office of Research, Demonstrations, and Statistics. Professional Standards Review Organization, 1979 Program Evaluation. HCFA Publication No. 03041. Washington, D.C.: U.S. Government Printing Office, 1980.

Call for Abstracts
American Public Health Association
1982 Annual Meeting

The Medical Care Section of the American Public Health Association is calling for abstracts of research dealing with the organization, financing, and utilization of health care. These abstracts must describe projects that will be completed by June 1982. The papers selected will be presented at the 110th Annual Meeting to be held November 15–18, 1982, in Montreal, Canada.

Abstracts must use the standard form appearing in the January and February 1982 issue of *The Nation's Health* or available from the contract below. All submissions to the Medical Care Section must be postmarked no later than March 12, 1982. Please submit to: Nelda McCall, Director, Health Policy Research, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025.

Appendix
Select Medical Resource Utilization and Cost Profiles

		Cohort			Utilizers	
Cohort	Resource Category	Total	Mean	Standard Deviation	Number	Mean
1	Confirmed, surgery elected (N = 305)					
	Physician visits	1,131	3.70	7.3	171	6.6
	Hospital stays	324	1.10	0.5	297	1.1
	Number of surgeries	395	1.30	0.7	305	1.3
	Total cost	$782,612	$2,566	2,788	305	$2,550
2	Confirmed, surgery not elected (N = 37)					
	Physician visits	75	2.00	4.6	17	4.3
	Hospital stays	1	0.03	0.2	1	1.0
	Number of surgeries*	1	0.03	0.2	1	1.0
	Total cost	$4,682	$127	342	17	$263
3	Not-confirmed, no surgery (N = 213)					
	Physician visits	552	2.60	7.0	92	5.9
	Hospital stays	12	0.60	0.2	11	1.0
	Number of surgeries*	4	0.02	0.1	4	1.0
	Total cost	$48,754	$229	1,227	99	$492
4	Not-confirmed, modified surgery (N = 44)					
	Physician visits	209	4.80	7.6	28	7.0
	Hospital stays	47	1.10	0.6	39	1.2
	Number of surgeries	82	1.90	1.0	43	1.9
	Total cost	$125,009	$2,841	1,831	44	$2,810
5	Not-confirmed, surgery elected after intervening medical treatment (N = 34)					
	Physician visits	188	5.50	5.7	29	6.5
	Hospital stays	41	1.20	0.4	34	1.2
	Number of surgeries	45	1.30	0.7	34	1.3
	Total cost	$119,751	$3,522	3,197	34	$3,484
6	Not-confirmed, surgery elected without intervening medical treatment (N = 51)					
	Physician visits	128	2.50	3.6	26	4.6
	Hospital stays	51	1.00	0.4	49	1.0
	Number of surgeries	62	1.20	0.7	51	1.2
	Total cost	$132,024	$2,589	3,526	51	$2,559
7	Not-confirmed, consultant's recommendation followed† (N = 291)					
	Physician visits	949	3.30	7.0	150	6.2
	Hospital stays	100	0.30	0.6	84	1.2
	Number of surgeries	131	0.50	0.9	81	1.6
	Total cost	$293,514	$1,009	2,102	177	$1,643

* Related surgeries.

† This group consists of cohorts 3, 4, and 5.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Impact of a Mandatory Second-Opinion Program on Medicaid Surgery Rates

SUZANNE GRISEZ MARTIN, M.P.A.,[*] MICHAEL SHWARTZ, PH.D.[†]
BERNADETTE J. WHALEN, B.S.,[‡] DEBORAH D'ARPA, M.B.A.,[*]
GRETA M. LJUNG, PH.D.,[†] JOHN HOLDEN THORNE, B.A.,[*]
AND ANNE E. MCKUSICK, B.A.[*]

The effect of the Massachusetts second-opinion program on the volume of elective surgery in the Medicaid population was assessed using two approaches: a study of the program experience and surgery decisions of 2,501 program referrals, and an analysis of Medicaid surgery rates before and after program implementation. Nonconfirmation rates, which averaged 14.5 per cent, varied by procedure from 4 per cent for cholecystectomy to 26 per cent for disc surgery. The patient's surgery decision was related to the outcome of the second-opinion consultation: 85.5 per cent of the confirmed patients had the originally proposed operation, as compared with 31 per cent of the nonconfirmed patients. In the year after program implementation, the program was associated with a 20 per cent reduction in the volume of those procedures covered by the program. The greatest percentage declines were for hysterectomies, meniscectomies, hemorrhoidectomies and tonsillectomies/adenoidectomies. The decline in surgery rates was attributed both to a direct effect on patients referred to the program and to a sentinel effect whereby fewer operations were proposed. We conclude that the mandatory second-opinion program in Massachusetts saved Medicaid $3 to $4 for every dollar spent.

Statement of the Problem

BETWEEN 1971 AND 1977, the rate of surgery performed in the United States increased by 27.5 per cent. By 1977, one surgical procedure was being performed for every 100 persons.[1] Whether this increase represents improved access to care or greater utilization of surgical treatment relative to medical treatment is under debate. But there is no doubt that this growth has contributed to the nation's steadily increasing health expenditures.

Various public and private health insurers are attempting to control surgery expenditures by encouraging or requiring their enrolled populations to obtain a second opinion on the need for proposed elective operations. Advocates of this approach argue that second-opinion programs are able to identify operations that can be deferred or not performed without jeopardiz-

---

From the Massachusetts Health Research Institute,[*] the School of Management, Boston University,[†] and the Massachusetts Department of Public Welfare.[‡]

This study was supported by a grant from the Health Care Financing Administration, U.S. Department of Health and Human Services (grant #18-P-97037/1-01) and with funds from the Commonwealth of Massachusetts. The authors would like to thank Richard Potts, Carol Greenfield and Joseph Blackburn for their contribution and the Health Care Foundations of Massachusetts for their cooperation.

Address for reprints: Suzanne Grisez Martin, 134 A Petty Rd., RD #2, Cranbury, NJ 08512.

0025-7079/82/0100/0021/$01.75 © J. B. Lippincott Co.

ing the patient's health. They assume that these programs save money by lowering the amount of elective surgery, both by a direct program effect on patients who decide against the operation after a negative second opinion and by a sentinel effect whereby fewer operations are actually proposed.[2] Second-opinion programs are also described as consumer oriented, since they increase the patient's role in treatment decisions by providing an additional opportunity to explore the advantages and risks of a particular operation.

In 1976 the Massachusetts legislature mandated the establishment of a second-opinion program for Medicaid recipients. This paper presents the results of a study of this program, including its effects on the number of operations performed on the Medicaid population and the consequent cost savings to Medicaid.[3]

Program Description

The Consultation Program for Elective Surgery (CPES) requires that certain Massachusetts Medicaid recipients proposed for any of eight elective procedures participate in a second-opinion program as a precondition for Medicaid reimbursement for the surgery. After completing program requirements, patients may choose to have the originally proposed operation even if the second-opinion consultants do not confirm the need for surgery, and physician and hospital claims will be reimbursed according to the Medicaid fee schedule.

The Department of Public Welfare is responsible for the design and implementation of the program. Medicaid recipients without other health insurance coverage, primarily children and women under the age of 65, must participate in the program if proposed for one of the following eight procedures: tonsillectomy/adenoidectomy; hysterectomy; cholecystectomy; submucous resection/rhinoplasty; hemorrhoidectomy; meniscectomy; excision and stripping of varicose veins; and disc surgery/spinal fusion. The Department contracts with the state's five health care foundations to administer the program. Each foundation covers an area corresponding to one of the state's five Professional Standard Review Organization (PSRO) regions.

The surgeon proposing the operation refers the patient to the foundation. When a patient is required to have a consultation, the foundation arranges an appointment with a board-certified/eligible surgeon in the appropriate surgical specialty. The consultant examines the patient and either does or does not confirm the need for surgery. The foundation is responsible for notifying the proposing surgeon, the hospital and the patient of the consultant's decision. If the first consultant does not confirm the need for surgery, it must also inform the patient of the requirement to receive another consultation if he/she wants to have the operation. The identity of the proposing surgeon is concealed on any information forwarded to the consultant, and the proposing surgeon is not told the name of the consultant. The consultant is not allowed to perform the operation.

The program was phased in over a 14-month period by region, beginning in March 1977. There are two different program models. Three regions use a "hands-on" model in which persons referred to the program are examined by a foundation-assigned surgeon who assesses the need for the proposed operation. The other two regions use a "desk-audit" model in which proposed patients are screened by the foundation using preset criteria developed by area physicians in consultation with the Department of Public Welfare. The foundation notifies those patients not meeting the criteria and their proposing surgeons that a hands-on consultation is required if the patient is still interested in having the proposed operation.

Methodology

Two distinct approaches were used to examine the impact of the CPES on the number of operations performed on the Massachusetts Medicaid population. The first was a study of the surgery decisions of all 2,501 persons referred to the program in the Bay State (Boston Metropolitan area) and Central Massachusetts (Worcester area) regions during a one-year period (May 1, 1977 to April 30, 1978 in the former area and October 1, 1977 to September 30, 1978 in the latter). These two areas cover 65 per cent of the eligible population, and both employ the hands-on model. They were selected for special study because they implemented the program the earliest, in March and August 1977, respectively.

Program files maintained by both foundations were used as the data source for information on the demographic characteristics and program experience of patients referred to the CPES. Consultant reports were reviewed to determine both the reasons for nonconfirmations and whether any conditions were imposed on the confirmations. The Medicaid billing system and hospital records were used to track each referred patient from the date of referral to CPES through December 31, 1978, to determine whether the patient had the originally proposed CPES operation, no surgery or a procedure related to the originally proposed operation. Related surgeries included diagnostic procedures (such as arthroscopies performed on patients proposed for meniscectomies) and alternative surgical procedures (such as dilations and curettages performed on women proposed for hysterectomies). Fifty-nine per cent of the patients were followed for one year or more; 36 per cent for six to 12 months, and 5 per cent for three to five months.

Summary information on the disposition of patients referred to the CPES in the Charles River, Western and Southeastern Massachusetts regions was extracted from monthly reports submitted to the Department by each foundation.

The second approach analyzed monthly surgery rates for each CPES procedure in each of the state's five regions over a 44-month period spanning CPES implementation. Physician claims submitted to Medicaid were used to determine the number of operations. A total of 27,000 claims was reviewed to ensure that only one claim for any single operation was counted and that the patient receiving the operation was covered by CPES. Monthly counts of the number of individuals eligible for the program were obtained from the Department's monthly summary of expenditure reports.

For each procedure in each region, 44 months of surgery rates were calculated from May 1975 (the earliest point at which surgery codes were fully implemented in the Medicaid billing system) to December 1978 (determined by project deadline). Monthly rates were used to ensure significant data points to estimate pre- and post-CPES levels of surgery. Because of the 14-month spread in implementation dates among the five regions, the number of months of pre- and post-CPES data varied by region. There were between 22 and 37 months of pre-CPES data and between seven and 22 months of post-CPES data.

Monthly surgery rates for each procedure in each region were analyzed to determine if a statistically significant change ($p < 0.05$) in rates occurred between the pre- and the post-CPES periods.[4] When monthly surgery rates were independent, the statistical significance of a change in the level of surgery was performed using an ordinary t-test, or, when low surgery volume suggested nonnormality, a Mann-Whitney U test. When autocorrelation in the rates was present, the statistical significance of a change in level was tested using the procedure proposed by Box and

Tiao.[5] Basically, this amounts to performing an ordinary t-test after allowing for correlation among the observations.

The analysis was complicated by the fact that Medicaid reduced reimbursement fees for surgery by 30 per cent in February 1976. The same approach used to assess the impact of CPES was used to analyze the impact of the fee cut. When the fee cut had a statistically significant impact on monthly surgery rates, data prior to the fee cut were eliminated before evaluating the impact of CPES.

When a statistically significant change was detected in monthly surgery rates after CPES implementation, two rival hypotheses were considered as alternative explanations. The first hypothesis was that the change was an artifact of a gradual declining trend in surgery rates that had begun prior to the program and continued after the program came into effect. This possibility was evaluated by analyzing the statistical significance of a trend in the pre-CPES surgery rate data.

The second hypothesis was that factors unrelated to CPES influenced surgery rates at the time CPES was implemented. For Bay State and Central Massachusetts, this hypothesis was analyzed by considering as a nonequivalent control group the three regions of the state that had not yet implemented a CPES program. The monthly surgery rates of this control group were examined to determine if a statistically significant drop in rates occurred concurrently with CPES implementation in Bay State and Central Massachusetts. Such a drop in the control group would indicate that factors other than CPES might have caused the detected changes in Bay State and Central Massachusetts.

Estimates of the number of operations not performed as a result of CPES, that is, forgone surgeries, were determined from the analyses of program referrals and surgery rates. The number of forgone surgeries calculated from the analysis of program referrals provided an estimate of the direct impact of CPES on persons proposed for surgery and referred to the program. The number of forgone surgeries calculated from the surgery rate analysis provided an estimate of the total impact of the program, that is, the direct impact on persons referred to the program, as well as any impact that results because fewer people were initially proposed for surgery (the sentinel effect). To estimate the sentinel effect of CPES, the direct program impact based on program participants was subtracted from the total impact based on the surgery rate analysis.

The cost-effectiveness analysis involved a comparison of program expenditures and program savings to Medicaid that resulted from the forgone surgeries. Program expenditures included administrative costs of the foundations and the Department of Public Welfare, fees paid to physician consultants and estimated Medicaid payments for diagnostic tests ordered during consultations. Estimates of gross savings to Medicaid were calculated by multiplying estimates of the number of forgone surgeries by the average Medicaid payment for each procedure. Average payments per procedure were calculated from a review of the Medicaid claims of 372 program participants that had CPES operations and hospital per diem rates for each PSRO based on information from the Massachusetts Rate Setting Commission. Surgery-related costs included average Medicaid payments to the hospital, surgeon, assistant surgeon and anesthesiologist; payment for drugs and diagnostic tests ordered during hospitalization; and follow-up visits to the surgeon of record.

Gross program savings were adjusted to account for costs of alternative treatment for patients forgoing the CPES surgery. These costs were determined from an analysis of Medicaid claims for those program participants in Bay State and Central Massachusetts who did not have the originally proposed surgery. Alternative surgical costs were estimated from actual Medicaid payments for related surgeries of program participants. Alternative medical treatment costs for each procedure were estimated by comparing Medicaid payments for a sample of patients having surgery, less the surgery costs, with Medicaid payments for a sample of patients forgoing surgery. Medicaid payments for a six-month period after the date of program referral for all services possibly related to the surgery recommendation were calculated for each sample, and the geometric means (the means after logarithmic transformation of the data) were compared.

Findings

Program Statistics

Bay State and Central Massachusetts. Of the 2,501 persons referred to the Consultation Program for Elective Surgery (CPES) in these regions, most were proposed for tonsillectomies/adenoidectomies (40 per cent of all referrals) and hysterectomies (24 per cent of all referrals). Patients proposed for each of the other six CPES procedures accounted for no more than 10 per cent of all referrals. One third of the referred patients were under age 15, and 85 per cent of those over age 15 were women.

Eighty-one per cent of the referred patients had a physician consultation; 10 per cent canceled from the program prior to a consultation and 9 per cent had the consultation requirement waived (see Table 1). The majority of cancellations were patients who failed to keep a scheduled appointment. Other patients canceled because they were inappropriately referred (they had other insurance or were proposed for non-CPES procedures) or were admitted as emergencies (and thus not required to participate).

Waivers from the consultation requirement were granted when the nurse-coordinator at the foundation accepted the proposing surgeon's request for a waiver because of medical necessity (patient in severe pain or already in hospital) or when the patient was 15–20 miles from the nearest consultant (fewer than 10 per cent of all waivers were granted for travel hardship). Medical waivers were granted most frequently for patients proposed for disc surgery/spinal fusion and for hysterectomies, primarily for patients with the diagnosis of cancer in situ.

TABLE 1. Distribution of 2,501 Persons Referred to the CPES by Proposed Procedure: Bay State and Central Massachusetts

Procedure	Number Having Consultation (%)	Number Canceling (%)	Number Granted Waiver (%)	Total Number Referred (%)
Tonsillectomy/adenoidectomy	910 (89.9)	92 (9.1)	10 (1.0)	1,012 (100)
Hysterectomy	425 (70.3)	62 (10.2)	118 (19.5)	605 (100)
Cholecystectomy	202 (79.2)	23 (9.0)	30 (11.8)	255 (100)
Submucous resection/ rhinoplasty	172 (84.3)	30 (14.7)	2 (1.0)	204 (100)
Disc surgery/spinal fusion	84 (62.2)	14 (10.4)	37 (27.4)	135 (100)
Meniscectomy	109 (82.6)	12 (9.1)	11 (8.3)	132 (100)
Excision and stripping of varicose veins	89 (85.6)	9 (8.6)	6 (5.8)	104 (100)
Hemorrhoidectomy	43 (79.6)	8 (14.8)	3 (5.6)	54 (100)
Total	2,034 (81.3)	250 (10.0)	217 (8.7)	2,501 (100)

The final decisions of the physician consultants on the 2,034 patients who received consultations are presented in Table 2. For 85.5 per cent of the patients, the first consultant confirmed the need for the proposed surgery. The confirmation rate exceeded 90 per cent for patients proposed for cholecystectomy, excision and stripping of varicose veins, and hemorrhoidectomy. Review of the consultation reports indicated that 3 per cent of the confirmations were stated conditionally, for example, confirmed *if* the patient lost weight, or *if* an arthroscopy indicated the need.

The first consultant did not confirm the need for the proposed surgery for 294 patients. Almost half of these (44 per cent) had a second consultation. Of the nonconfirmed patients who sought a second consultation, 65 per cent were confirmed for CPES surgery.

The highest nonconfirmation rates were for patients proposed for disc surgery/spinal fusion (26 per cent), hysterectomy (23 per cent) and meniscectomy (24 per cent). The most frequent reason stated for nonconfirmation was that the proposed operation was "not medically necessary." Approximately 35 per cent of the nonconfirmations included qualifying statements by the consultant as to why the proposed procedure was not the optimum treatment at the time of the consultation. These reasons included recommendations that medical therapy or less extensive surgery be explored first; the need for more testing and/or a longer observation period; and statements that the patient was a poor risk for surgery.

Ten per cent of the patients received one or more diagnostic tests from the first consultant. The most frequently ordered tests were blood tests, Pap smears, anoscopies and cultures.

Charles River, Western and Southeastern Massachusetts. The program disposition of patients referred to the CPES in the Charles River, Western and Southeastern

TABLE 2. Distribution of 2,034 Persons Receiving Physician Consultations Through the CPES by Proposed Procedure and Outcome of the Consultation: Bay State and Central Massachusetts

	Distribution of Patients by Outcome of the First Consultation		Distribution of Patients Not Confirmed by the First Consultant		
Procedure	Confirmed by First Consultant No. (%)	Not Confirmed by First Consultant No. (%)	No Second Consultation No. (%)	Confirmed by Second Consultant No. (%)	Not Confirmed by Second Consultant No. (%)
Tonsillectomy/ adenoidectomy	814 (89.5)	96 (10.5)	68 (70.8)	21 (21.9)	7 (7.3)
Hysterectomy	326 (76.7)	99 (23.3)	47 (47.5)	35 (35.3)	17 (17.2)
Cholecystectomy	193 (95.5)	9 (4.5)	4 (44.4)	4 (44.4)	1 (11.1)
Submucous resection/ rhinoplasty	142 (82.6)	30 (17.4)	13 (43.3)	9 (30.0)	8 (26.7)
Disc Surgery/ spinal fusion	62 (73.8)	22 (26.2)	7 (31.8)	10 (45.5)	5 (22.7)
Meniscectomy	83 (76.1)	26 (23.9)	18 (69.2)	3 (11.5)	5 (19.2)
Varicose veins	81 (91.0)	8 (9.0)	5 (62.5)	2 (25.0)	1 (12.5)
Hemorrhoidectomy	39 (90.7)	4 (9.3)	2 (50.0)	1 (25.0)	1 (25.0)
Total	1,740 (85.5)	294 (14.5)	164 (55.8)	85 (28.9)	45 (15.3)

TABLE 3. Distribution of 2,363 Patients Referred to the CPES: Charles River, Western Massachusetts and Southeastern Massachusetts

Region	No. Referred	No. (%) of Referrals Meeting Criteria	No. Having One Consultation	% Confirmation Rate: First Consultation	No. Having Two Consultations
Charles River	257	not applicable	222	87.8	13
Western Massachusetts	942	732 (77.2)	143	74.1	25
Southeastern Massachusetts	1,164	1,080 (92.8)	74	91.9	4

Massachusetts regions during the first year of operation in each site is presented in Table 3. The Charles River statistics are similar to those of Bay State and Central Massachusetts, but of smaller volume since the area encompasses only 3 per cent of the eligible CPES population. In Western and Southeastern Massachusetts, where the desk-audit model is used, 2,106 patients were screened using the preset criteria. Seventy-seven per cent of Western referrals and 93 per cent of Southeastern referrals met the criteria and were therefore not required to have a second-opinion consultation. The confirmation rate at the first consultation for those patients not meeting the criteria was 74 per cent in Western (n = 143) and 92 per cent (n = 74) in Southeastern Massachusetts. Only 29 patients in both areas had two second-opinion consultations.

Effect of the CPES on Number of Surgeries Performed

Surgery Decisions of Bay State and Central Massachusetts Patients. Seventy-two per cent of the 2,501 patients referred to the CPES had the originally proposed procedure; 22 per cent had no surgery, and 4 per cent had a related surgery. The

TABLE 4. Surgery Decisions of 2,501 Persons Referred to the CPES by Proposed Procedure: Bay State and Central Massachusetts

Procedure	No Surgery No. (%)	CPES Surgery No. (%)	Related Surgery No. (%)	Outcome Unknown No. (%)	Total No. (%)
Tonsillectomy/adenoidectomy	211 (21.8)	786 (77.7)	2 (0.2)	3 (0.3)	1,012 (100.0)
Hysterectomy	118 (19.5)	423 (69.9)	53 (8.8)	11 (1.8)	605 (100.0)
Cholecystectomy	36 (14.1)	211 (82.8)	1 (0.4)	7 (2.7)	255 (100.0)
Submucous resection rhinoplasty	64 (31.4)	128 (62.7)	1 (0.5)	11 (5.4)	204 (100.0)
Disc surgery/spinal fusion	31 (23.0)	84 (62.2)	17 (12.6)	3 (2.2)	135 (100.0)
Meniscectomy	34 (25.8)	67 (50.8)	25 (18.9)	6 (4.5)	132 (100.0)
Varicose veins	36 (34.6)	67 (64.4)	0 (0.0)	1 (1.0)	104 (100.0)
Hemorrhoidectomy	15 (27.7)	34 (63.0)	5 (9.3)	0 (0.0)	54 (100.0)
Total	555 (22.2)	1,800 (71.9)	104 (4.3)	42 (1.7)	2,501 (100.0)

surgery status of 2 per cent of the patients is unknown because they were ineligible for Medicaid at the time of follow-up, the Medicaid number was unavailable, or there was an unlisted prodecure code during a hospitalization. Ninety-two per cent of the patients having surgery did so within three months of the date on which their proposing surgeon referred them to the program. Patients proposed for cholecystectomy, tonsillectomy/adenoidectomy and hysterectomy were most likely to have the originally proposed operation (see Table 4). Related surgery as an alternative to the CPES surgery was most frequently performed on patients proposed for meniscectomy (mostly arthroscopies), disc surgery (mostly myelograms) and hysterectomy (mostly dilations and curettages).

There was a strong relationship between the first consultant's opinion and the patient's surgery decision: 85.5 per cent of patients whose need for a CPES operation was confirmed by the first consultant had the surgery, as compared with 31 per cent of those not confirmed (see Table 5). Of patients not confirmed by the first consultant, those least likely to have the CPES operation were those deciding against a second consultation (only 7 per cent of these patients had the originally proposed operation, in apparent disregard of CPES regulations). Forty per cent of the patients not confirmed by either the first or second consultant had the operation anyway, as did 74 per cent of those who received a nonconfirmation from the first consultant and a confirmation from the second consultant. Seventy-eight per cent of the patients waived from the consultation and 20 per cent of those patients who canceled also had the originally proposed operation.

To identify the possible range of program impact, project staff made both minimum and maximum estimates of the number of patients referred to the CPES who did *not* undergo the proposed operation as a result of the program. The minimum estimate assumed that only a portion of those patients who received nonconfirmations were deterred from surgery by the program. The number of nonconfirmed patients who would have decided against surgery without the pro-

TABLE 5. Distribution of 2,501 Persons Referred to the CPES by Program Disposition and Surgery Decision: Bay State and Central Massachusetts

Program Disposition	No Surgery No. (%)	CPES Surgery No. (%)	Related Surgery No. (%)	Unknown Surgery No. (%)	Total No. (%)
Confirmed first consultation	180 (10.3)	1,487 (85.5)	53 (3.0)	21 (1.2)	1,740 (100.0)
Not confirmed, first consultation, no second consultation	137 (83.5)	11 (6.7)	12 (7.3)	4 (2.4)	164 (100.0)
Not confirmed first consultation, confirmed second	15 (17.6)	63 (74.1)	5 (5.9)	2 (2.4)	85 (100.0)
Not confirmed first consultation, not confirmed second	23 (51.1)	18 (40.0)	2 (4.4)	2 (4.4)	45 (100.0)
Granted waivers	24 (11.1)	170 (78.3)	17 (7.8)	6 (2.8)	217 (100.0)
Canceled	176 (70.4)	51 (20.4)	16 (70.4)	7 (2.8)	250 (100.0)
Total	555 (22.2)	1,800 (72.0)	104 (4.1)	42 (1.7)	2,501 (100.0)

gram was estimated from the percentage of confirmed patients who elected not to have surgery. This number was subtracted from the number of nonconfirmed patients who did not have CPES operations to establish the minimum estimate. The maximum estimate assumed that the patients who decided against surgery because of the program included both those who canceled from the program by not appearing for their consultation and those who had a consultation and did not have the surgery. These patients might have decided against surgery because of information provided by the consultant or due to the additional time the program gave them to consider the operation. CPES operations might not be performed on patients canceling without completing program requirements because provider claims would be subject to disallowance by Medicaid.

The minimum estimate of program impact on program participants in Bay State and Central Massachusetts is 177 forgone surgeries, representing 27 per cent of all patients referred to the program in these regions who did not have the originally proposed surgery. The maximum estimate of forgone surgeries is 577, or 85 per cent of all program participants not having the originally proposed surgery in Bay State and Central Massachusetts. This maximum estimate includes 232 confirmed patients, 194 nonconfirmed patients and 131 patients who failed to keep scheduled appointments. The number of forgone surgeries by procedure from this analysis of program participants is found in Table 9.

CPES Impact on Medicaid Surgery Rates

Table 6 presents the mean monthly surgery rates per 100,000 population by region for each procedure for the periods before and after CPES implementation. The table also indicates whether the change in surgery rates after CPES implementation is statistically significant. (In the Appendix, several graphs of the time series in Bay State and Central Massachusetts are presented and discussed).

Minimum and maximum estimates were made of the magnitude of CPES impact on Medicaid surgery rates. The minimum estimate was based only on data from the Bay State and Central Massachusetts regions. At least 15 months of post-CPES surgery rates were available for both of these regions, and the possibility that other factors caused the surgery rate decline after CPES implementation could be considered by comparison with the nonequivalent control group.

Three criteria were used to assess the effect of CPES on surgery rates in Bay State and Central Massachusetts: 1) whether there was a statistically significant drop in the mean rate of surgery after CPES implementation; 2) whether there was evidence of a statistically significant decline in the rate of surgery prior to CPES (thus suggesting that the rate drop might be an artifact of a declining trend); and 3) whether there was a comparable drop in mean surgery rates in the control group (suggesting that factors other than CPES were responsible for the observed decline). Procedures that met all three criteria were included in the minimum estimate of CPES impact.

The maximum estimate was composed of all procedures in any region in which a statistically significant drop in the mean level of surgery rates occurred concurrently with program implementation and in which no evidence of a trend prior to CPES was found. This statewide estimate should be considered preliminary since less than 12 months of post-CPES surgery rates were available for three of five regions.

The minimum estimate attributes 603 fewer surgeries in the Bay State and Central Massachusetts regions to the CPES in the year after program implementation, including 391 fewer tonsillectomies/adenoidectomies and 135 fewer hysterec-

TABLE 6. Average Monthly Surgery Rates Per 100,000 Medicaid Population in the Pre-CPES and Post-CPES Period By Procedure and Region*

Procedure	Bay State		Central		Western		Southeastern		Charles River	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Tonsillectomy/adenoidectomy	30.77	18.25†	64.11	48.97†	99.98	49.04†	86.84	49.25†	75.83	55.69
Hysterectomy	17.10	13.10†	15.41	11.73†	28.61	19.67†	22.98	16.51†	19.70	9.49†
Cholecystectomy	9.01	9.07	15.79	13.20	12.16	12.14	15.85	10.37†	8.93	16.82
Submucous resection/rhinoplasty	2.65	2.34	13.68	7.44†	9.34	5.74†	16.28	15.43	12.63	10.58
Disc surgery/spinal fusion	4.35	4.14	6.80	5.64	5.55	5.46	4.68	4.13	5.25	5.26
Meniscectomy	3.32	2.46†	2.82	3.86	2.13	1.95	3.88	2.13†	2.70	5.22
Varicose veins	3.32	2.39†	4.23	3.39	4.65	3.23	3.30	1.70	4.88	3.15
Hemorrhoidectomy	2.76	2.11†	5.10	3.40†	4.51	3.32	3.38	1.90†	3.97	2.15

* Pre-fee cut data were eliminated for those cases in which fee cut had a statistically significant effect on monthly surgery rates. The Pre- and Post-CPES periods in each region are as follows:

Bay State	1–22,	23–44
Central	1–29,	30–44
Western	1–34,	35–44
Southeastern	1–37,	38–44
Charles River	1–37,	38–44

(Month 1 corresponds to May 1975; month 44 corresponds to December, 1978.)

† Indicates that change in level is statistically significant ($p < 0.05$).

tomies (see Table 7). These 603 operations represent a 20 per cent decline in total CPES operations in these two regions, as compared with the year before implementation. The greatest percentage declines were for tonsillectomies/adenoidectomies (28 per cent), hysterectomies (23 per cent) and meniscetomies (22 per cent).

The maximum estimate of CPES impact attributes 1,139 forgone surgeries to CPES statewide, representing a 20 per cent decline in CPES operations statewide in the year after implementation, as compared with the preceding year. Hysterectomies showed the greatest decline (26 per cent), followed by hemorrhoidectomies (23 per cent), meniscectomies (23 per cent) and tonsillectomies/adenoidectomies (22 per cent). The only procedure with no evidence of a statistically significant decline in surgery rates in any region was disc surgery/spinal fusion. The percentage decline in CPES operations varied by region from a 7–8 per cent decline in two regions to a greater than 20 per cent decline in three.

The Sentinel Effect

The sentinel effect, which refers to the effect of the program on the amount of proposed surgery, was measured by subtracting the direct program effect (estimated from the analysis of program participants) from the total program effect (estimated from the surgery rate analysis). In Bay State and Central Massachusetts, for which there are minimum and maximum estimates of both the direct and total impact, there are corresponding minimum and maximum estimates of the sentinel effect. The minimum estimate of the sentinel effect was calculated by subtracting the maximum estimate of direct impact from the minimum estimate of total impact; the maximum estimate of the sentinel effect was calculated by subtracting the minimum estimate of direct impact from the maximum estimate of total impact. Table 8 presents the minimum and maximum estimates of the sentinel effect for Bay State and Central Massachusetts. Tonsillectomies/adenoidectomies show

TABLE 7. Minimum and Maximum Estimates of the Number of Forgone CPES Operations Attributable to the CPES by Procedure Based on Surgery Rate Analysis

	Minimum Estimate (Bay State and Central)			Maximum Estimate (Statewide)		
Procedure	No. Forgone Operations	No. Operations 12 Months Prior to CPES	% Change	No. Forgone Operations	No. Operations 12 Months Prior to CPES	% Change
Tonsillectomy/ adenoidectomy	391	1,385	28	641	2,878	22
Hysterectomy	135	595	23	276	1,054	26
Cholecystectomy	0	374	0	45	622	7
Submucous resection/ rhinoplasty	33	170	19	71	429	17
Disc surgery/spinal fusion	0	138	0	0	254	0
Meniscectomy	25	114	22	39	172	23
Varicose veins	0	126	0	27	196	14
Hemorrhoidectomy	19	113	17	40	175	23
Total	603	3,015	20	1,139	5,780	20

the most evidence of a sentinel effect. There is some sentinel effect for all procedures except cholecystectomies and disc surgery, but, with the exception of hysterectomies, the number of operations involved is small.

In Western and Southeastern Massachusetts, where the desk-audit model is used, the maximum direct program effect was estimated as the number of cases that did not meet the preset criteria. A minimum sentinel effect was calculated by subtracting this maximum estimate of direct effect from the total effect based on the estimates from the surgery rate analysis. In Southeastern Massachusetts the sentinel effect seems to account for most of the impact of CPES on tonsillectomies/adenoidectomies, cholecystectomies and hysterectomies, since very few patients did not meet the criteria and fewer still received nonconfirming second opinions. In Western Massachusetts there was evidence of a sentinel effect only for submucous resection.

Projected Savings of CPES

Medicaid expenditures for CPES during the first year of program operation in all regions were $293,800, with 77 per cent of the budget spent for administration, 22.5 per cent for physician fees and 0.5 per cent for diagnostic tests ordered by consultants and billed to Medicaid.

Costs of alternative treatment for those patients forgoing CPES operations were grouped into costs of alternative surgical treatment and costs of alternative medical treatment. Costs for alternative surgical treatment ranged by procedure from 0 per cent of unadjusted gross savings for varicose veins to 40 per cent for hemorrhoidectomy. On average, the cost of alternative surgical treatment represented 8 per cent of gross savings.

No adjustment was made for alternative medical treatment costs because there was no statistically significant difference between the average Medicaid payments for patients referred to CPES who had the

TABLE 8. Minimum and Maximum Estimates of the Number of Forgone Operations Attributed to the Sentinel Effect by Region and Procedure

	Bay State		Central Massachusetts		Western Massachusetts	Southeastern Massachusetts
Procedure	Minimum No. (%)*	Maximum No. (%)*	Minimum No. (%)*	Maximum No. (%)*	Minimum No. (%)*	Minimum No. (%)*
Tonsillectomy/adenoidectomy	231 (59)	350 (90)	0	55 (67)	—	166 (99)
Hemorrhoidectomy	3 (16)	18 (95)	0	9 (100)	—	0
Cholecystectomy	—	—	—	—	—	42 (93)
Meniscectomy	0	8 (32)	—	—	—	9 (64)
Submucous resection/rhinoplasty	0	0	7 (21)	26 (79)	24 (83)	—
Varicose veins	0	23 (85)	—	—	—	—
Hysterectomy	2 (2)	72 (63)	6 (30)	12 (60)	0	15 (25)
Disc surgery/spinal fusion	—	—	—	—	—	—
Total	236 (40)	471 (80)	13 (9)	102 (71)	24 (24)	230 (79)

The symbol — indicates that CPES had no statistically significant impact on surgery rates; 0 indicates that CPES had an impact, but that none of the impact is attributed to the sentinel effect.

* Numbers in parentheses denote the percentage of the total program impact that is due to the sentinel effect.

CPES operation (less payment for the surgery) and for patients forgoing the CPES operation.

The statewide estimate of net program savings for the year following CPES implementation in all regions is $856,506, for a 1:3.9 cost–benefit ratio. This estimate was derived from the 1,139 forgone CPES procedures statewide in the year after program initiation.

There are four estimates of program savings for the Bay State and Central Massachusetts regions, each derived from one of the four different estimates of forgone surgeries. The number of forgone surgeries and corresponding net program savings by procedure for each of these four estimates for the two regions is presented in Table 9. Savings estimates for these two regions range from $96,184, for a cost–benefit ratio of 1:1.5, to $657,265, for a cost–benefit ratio of 1:4.7.

Discussion

Gertman et al. have reported that the CPES "may produce only modest direct savings in Medicaid expenditures."[6] This conclusion was based on an analysis of the program disposition of CPES referrals to the Bay State Health Care Foundation and a follow-up of those patients referred for hysterectomies. Program benefits were calculated by comparing the estimated costs associated with surgery events of women who received negative second opinions with the estimated costs for these women if their surgery experience paralleled that of women with positive opinions. The authors estimate a 2.27 benefit–cost ratio for hysterectomies in Bay State. Using similar assumptions in a "more speculative" analysis, for which the surgery decisions of all referred patients were not known, they estimate an overall benefit–cost ratio of 1.1 for CPES in Bay State.

Our minimum estimate of program savings based on program referrals in Bay State, which also assumes a CPES effect only on some patients not confirmed for the proposed surgery, also predicts somewhat modest savings to Medicaid (a benefit–cost ratio of 1.83, based on follow-up of all Bay State referrals and including costs of alternative treatment). However, other aspects of our analysis have lead us to conclude that this estimate is unrealistically low because it excludes any sentinel effect and any direct impact on referrals other than those with negative opinions on the need for the proposed operation.

The only other published data on mandatory second-opinion programs come from several security welfare funds in the greater New York City area. McCarthy has analyzed these programs through observation of program participants and has reported on referral volume, consultation outcomes and surgery status. In 1978, McCarthy reported an overall nonconfirmation rate of 18.5 per cent on the 5,025 second-opinion consultations provided in the mandatory program since 1972.[7] This is slightly higher than the 14.5 per cent nonconfirmation rate for 2,501 consultations in the CPES during its first year in Bay State and Central Massachusetts. In both programs approximately two thirds of the patients not confirmed did not have the originally proposed surgery. McCarthy reports that 79.4 per cent of the nonconfirmed/no surgery patients interviewed stated that they had elected against surgery because the consultant had advised against it.

The assessment of the Massachusetts second-opinion program presented here includes both an analysis of program participants and an analysis of the target population's surgery rates before and after program implementation. These two approaches result in different estimates of program impact. The advantage of the analysis of program participants is that the number of forgone surgeries is based on a direct count of persons referred to CPES. Its disadvantages are the difficulty of de-

TABLE 9. Minimum and Maximum Estimates of Net Program Savings* in the Bay State and Central Massachusetts Regions for the Year Following Program Implementation

Procedure	Estimated Net Savings Based on Analysis of Program Participants				Estimated Net Savings Based on Surgery Rate Analysis			
	Minimum		Maximum		Minimum		Maximum	
	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)	No. Forgone Operations	Net Savings ($)
Tonsillectomy/ adenoidectomy	68	-37,973	206	38,645	391	153,420	473	196,945
Hysterectomy	51	69,660	127	249,769	135	265,633	135	265,633
Cholecystectomy	4	-4,761	32	62,311	0	-15,805	0	-15,805
Submucous resection/ rhinoplasty	17	3,853	59	45,972	33	13,289	42	23,874
Disc surgery/ spinal fusion	13	44,390	35	141,520	0	-10,093	0	-10,093
Meniscectomy	18	21,337	51	64,670	25	27,834	25	27,834
Varicose veins	5	1,496	30	40,078	0	-6,281	27	37,301
Hemorrhoidectomy	1	-1,818	17	14,300	19	15,248	28	22,187
Total	177	96,184	557	657,265	603	443,245	730	547,876
Cost-benefit ratio†		1:1.5		1:4.7		1:3.5		1:4.1

* Net program savings equal gross savings less estimated payments for alternative treatment and program costs.

† Cost-benefit ratio equals gross savings less estimated payments for alternative treatment divided by program costs. For Bay State and Central Massachusetts, the program costs were $178,747.

termining which program participants decided against surgery because of the program and the impossibility of measuring the sentinel effect (fewer operations proposed because of the program).

The difficulty of assessing the impact of a second-opinion program based solely on simple confirmation or nonconfirmation rates was highlighted in Paris' report on the voluntary second-opinion program for New York City's employees.[8] Paris reported that patients often received much more information than a mere confirmation/nonconfirmation on the need for the proposed operation from the surgeon rendering the second opinion. In a review of consultant reports he found a large "gray area" involving 22 per cent of the cases, in which the consultant supplemented the negative/positive opinion with "qualifications." These "qualifications" included such statements as: "observe patient further." and "additional diagnostic information or evaluation on the need for the proposed surgery required." The statement written in the records probably reflects more extensive conversations between the consultant and the patient on the advantages and disadvantages of various treatment options. This additional communication may explain why some nonconfirmed cases decided to have the originally proposed procedure and why some confirmed patients decided not to. For this reason, nonconfirmation rates are of limited value in predicting which program participants are affected by the program.

The advantage of the analysis of surgery rates we have performed is that it focuses on the entire population eligible to participate in CPES and can thus measure both direct program impact and the sentinel effect without making any assumptions about which program participants were affected by the program. Furthermore, by taking into consideration pre- as well as post-CPES rates, two possible errors in the analysis are avoided: 1) attributing a decline in rates to CPES when the decline represents a continuation of preprogram trends and 2) interpreting a constant level of surgery after implementation as indicating no program effect on the surgery level, when it in fact represents a statistically significant decline from the pre-CPES baseline period.

The surgery rate analysis in Bay State and Central Massachusetts was strengthened by using a nonequivalent control group to evaluate whether factors other than CPES may have been responsible for any observed drop in surgery rates. The use of nonequivalent control groups to improve interrupted time series designs has been strongly recommended and applied in a variety of situations.[4,9,10] However, one should attempt to identify sources of bias that might result from nonequivalency and lack of randomization. An important source of nonequivalency in this case was surgery rates that were initially higher in the control area than in Central Massachusetts, and particularly in Bay State. Because of regression artifacts that might tend to lower surgery rates in the control area, there is a bias toward interpreting similar declines in Bay State and Central Massachusetts to something other than CPES, when in fact they may have been caused by CPES. Despite this potential bias, with one exception, there was no drop in surgery rates in the control area concurrent with CPES implementation in Bay State and Central Massachusetts.

One factor that has been cited as an alternative explanation for the decline in surgery rates is the PSRO program. As the PSRO program was at least as completely implemented in the control area as in Bay State and Central Massachusetts when CPES was implemented, if the PSRO program caused the decline in these regions, one would expect to observe a similar decline in the control (assuming that the

PSRO impact on surgery rates is similar statewide). However, as stated earlier, in only one case out of 16 possibilities did surgery rates in the control area decline when CPES was implemented in Bay State or Central Massachusetts.

A further illustration of the unlikelihood of the PSRO program causing the surgery rate decline that we have ascribed to CPES can be seen in Southeastern Massachusetts, where most of the CPES impact has been attributed to the sentinel effect. As Figure 1 shows, there was a drop in surgery rates for all procedures combined in month 15 (June 1976) and then a leveling off until month 37 (May 1978), when the CPES was implemented. The first drop *might* be accounted for by the PSRO program, which was implemented between April 1976 and July 1977, or by the February 1976 Medicaid fee cut. There is *no* reason to suspect, however, that the PSRO program caused the abrupt drop in rates following CPES implementation, when the PSRO program had been in effect for 10 months. Furthermore, the PSRO program is more likely to have a gradual than an abrupt impact.[11]

The surgery rate analysis is better suited for the detection of abrupt, rather than gradual, changes. Graphs of the time series indicate that when CPES did affect surgery rates, there was an immediate drop in the level of surgery after program implementation. In no case did post-CPES surgery rates exhibit a tendency to return to pre-CPES levels. However, post-CPES surgery rates were tracked for less than two years in all areas of the state. Thus, the longer-term impact of the program on surgery volume cannot be assessed by the project's data.

The approach used to assess the impact of the fee cut on surgery rates was to eliminate the pre-fee cut data only if a statistically significant impact was observed. However, since the impact of the fee cut might not have been detected because of the low power of the statistical tests, and since this approach might thus bias results, the analysis was performed with pre-fee cut data eliminated in all cases. There was only a minimal effect on our conclusions:

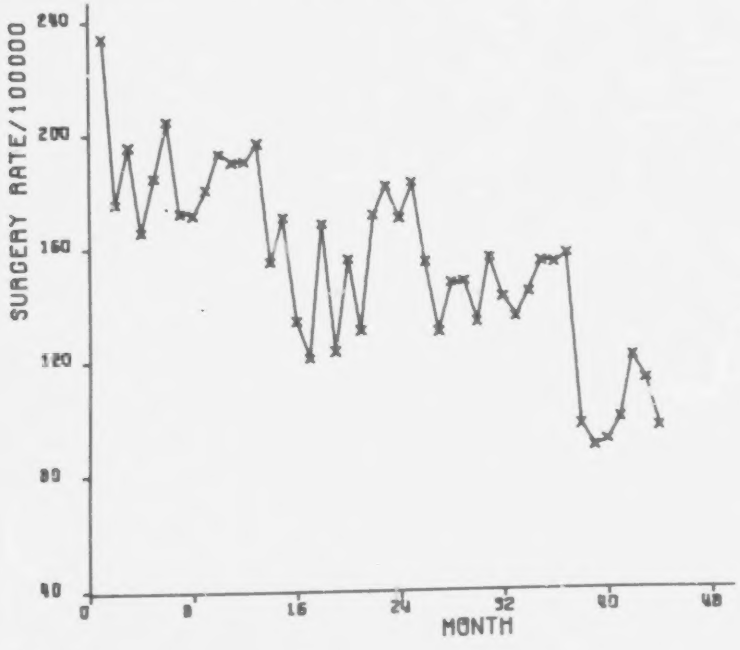


FIG. 1. Monthly rate of all CPES procedures in Southern Massachusetts.

the minimum cost–benefit ratio declined from 1:3.5 to 1:3.4; the maximum from 1:4.1 to 1:4.0; estimated statewide savings declined from $858,506 to $853,323.

Several points relative to the validity of the surgery rate data should be mentioned. The denominator data are based on monthly figures from routine Department of Public Welfare reports, and while there were monthly fluctuations, there were no significant differences in the age, sex or category-of-assistance distribution of the population over the 1976–1978 period or across the PSRO regions.

Numerator data are based on the physician claims submitted to Medicaid. This might be expected to result in some undercounting of the number of operations performed if there were no claims submitted for operations done by residents and if some physicians did not bill Medicaid because of low fees. In tracking which program participants had surgery, project staff compared hospital records with Medicaid files and found only 77 cases out of 1,800 (less than 5 per cent) in which a CPES operation was done without a corresponding physician bill in the Medicaid files. There is, therefore, a minimal amount of undercounting in the number of operations. However, only if a change in undercounting occurred once the CPES began would there be a bias in the analysis of surgery rates that has been presented.

There are two limitations of the surgery rate analysis. First, the approach may underestimate program impact for some procedures because the variability in monthly surgery rates made it unlikely that changes of less than 20 per cent would be identified as statistically significant. Second, the approach may overestimate the magnitude of CPES impact for other procedures, because if a statistically significant drop in surgery rates was detected, the entire amount of the drop was attributed to CPES, when other factors might have accounted for some of the decline.

The combination of both approaches provides insight into the effect of the CPES on different procedures. For example, despite the fact that there was no statistically significant drop in the rate of disc surgeries/spinal fusions in any region, the high nonconfirmation rate for this procedure suggests a possible program impact that was too small to be detected by the surgery rate analysis. Because of the expense of this procedure, a small impact on surgery rates would generate large dollar savings. For hysterectomies and meniscectomies, there were both a high rate of nonconfirmation and a large impact of the program on surgery rates. Tonsillectomies/adenoidectomies and hemorrhoidectomies had only a moderate nonconfirmation rate, but there was a large impact on surgery rates. This demonstrates that the nonconfirmation rate alone, which indicates only direct program effect, may not be a good indicator of the total impact of the program.

The effect of the program on surgery rates varied by region. Two of the regions using the hands-on model had a decline in surgery rates of over 20 per cent. The remaining region had a decline of slightly less than 10 per cent. However, in this region, which includes only 3 per cent of the eligible population, because of a short follow-up period, small volume and large variability in surgery rates, it was difficult to assess accurately the impact of the CPES. The short post-CPES period in the two regions with the desk-audit model permitted only a preliminary assessment of this model. One of these regions had a decline of more than 20 per cent in surgery rates following CPES; the other had a decline of just under 10 per cent. Since few people actually received consultations under the desk-audit model, program impact in these regions was attributed to a sentinel effect. It is interesting that the effect of the program on surgery rates did not seem related to the pre-CPES level of surgery. The area that had the lowest pre-CPES surgery rates had the largest decline (Bay State); the area with the second highest pre-CPES surgery rates had the second largest decline (Southeastern Massachusetts).

It is difficult to interpret the effect of CPES on surgery rates for each procedure statewide because of the multiple interventions: the fee cut plus the four different times at which CPES was phased in. However, it is interesting to compare the percentage decline in Medicaid surgery rates statewide by procedure with declines over the same period in surgery rates in the Northeastern region. As highlighted in Table 10, with the exception of disc surgery/spinal fusion, the decline in surgery rates in the Massachusetts Medicaid population was much larger than that in the Northeastern region.

TABLE 10. Yearly Percentage Change in Northeast Regional Surgery Rates (1975–1978)* Compared With Change in Massachusetts Medicaid Surgery Rates

Procedure	Regional Change (%)†	Massachusetts Medicaid Change (%)†
Tonsillectomy/adenoidectomy	−11	−25
Hysterectomy	−8	−16
Cholecystectomy	−1	−7
Submucous resection/rhinoplasty	+1	−8
Disc surgery/spinal fusion	−3	−1
Meniscectomy	−3	−12
Varicose veins	−5	−19
Hemorrhoidectomy	−16	−26

* From Vital and Health Statistics, Data from the National Health Survey, Series 13, no. 34 (Surgical Operations in Short Stay Hospitals: United States, 1975); no. 37 (Utilization of Short Stay Hospitals, Annual Summary for the U.S., 1977); no. 46 (Utilization of Short Stay Hospitals, Annual Summary for the U.S., 1978); and personal communication, National Center for Health Statistics.

† The regional change was calculated from the coefficient of the line fit by ordinary least squares to the logarithm of the four yearly rates from 1975–1978; the Massachusetts change was calculated from the coefficient of the line fit by ordinary least squares to the logarithm of the monthly surgery rates in the post-fee cut period.

The minimum and maximum estimates based on the two approaches can be used to focus attention on the effect of different assumptions on conclusions regarding CPES impact. Estimates of net CPES savings in Bay State and Central Massachusetts range from $96,184 to $657,256. Project staff believe that the lowest estimate, $96,184, underrepresents CPES impact because it omits any sentinel effect, any effect on patients who do not have surgery after missing a scheduled appointment and who may be deterred by bureaucratic barriers, and any effect on patients confirmed for the proposed operation who may defer or forgo surgery because of information provided by the consultant or as a result of additional time to consider the decision. The maximum estimate, $657,256, which attributes all decisions against surgery of confirmed and canceling patients to the program, but also omits any sentinel effect, may overstate CPES impact. Actual program savings for these two areas are probably between $443,245 and $547,876. The $443,245 figure translates into a cost–benefit ratio of 1:3.5—a realistic estimate of CPES savings for the two regions with the most available data for this analysis.

While our savings estimates consider payments for alternative medical treatments, the approach used is admittedly rough because there was no way to identify which Medicaid claims were related to a forgone CPES procedure. If alternative medical treatment costs were small, they would not have been detected; however, they would therefore not alter the net savings figures reported to any great extent. While a longer follow-up on medical treatment costs would have been interesting, six months is reasonable, given McCarthy's finding that there was no increase in the number of patients having alternative medical treatment between the six- and 12- month follow-up periods.

It should be reemphasized that reported savings are only from the perspective of Medicaid and do not address any costs or benefits to patients, providers, other insurers or to society. Further, the project did not address important quality-of-care issues associated with having or forgoing any of the CPES procedures.

Conclusions

The Massachusetts second-opinion program was associated with a 20 per cent reduction in the volume of the elective procedures covered by the program. The program was associated with a statistically significant decline in surgery rates for seven of the eight procedures in at least one of the five PSRO regions. The greatest declines were for hysterectomies, hemorrhoidectomies, meniscectomies and tonsillectomies/adenoidectomies. The only procedure with no significant decline was disc surgery/spinal fusion.

The reduction in the level of surgery generated savings to the Medicaid program that resulted in a return of $3–$4 for every dollar spent. Thus, the Massachusetts Medicaid program has realized substantial savings from its second-opinion program.

Future assessments of second-opinion programs should attempt to examine the effect of second opinions on quality of care, as well as the programs' value in providing additional information to patients to improve their ability to evaluate proposed operations.

References

1. American College of Surgeons. Socioeconomic factbook for surgery 1979. Haug JN, Kuntzman K (eds). Chicago: American College of Surgeons, 1979:35.

2. Grafe WR, McSherry CK, Finkel ML, et al. The elective surgery second opinion program. Ann Surg 1978;188:323.

3. Martin SG, Shwartz M, Cooper DD, et al. The effect of a mandatory second opinion program on medicaid surgery rates: an analysis of the Massachusetts Consultation Program for Elective Surgery. Baltimore: Health Care Financing Administration, DHHS, 1980.

4. Glass GV, Willson VL, Gottman JM. Design and analysis of time series experiments. Boulder: Colorado University Press, 1975.

5. Box GEP, Tiao GC. A change in level of a non-stationary time series. Biometrika 1965;52:181.

6. Gertman PM, Stackpole DA, Levenson DK, et al. Second opinions for elective surgery. N Engl J Med 1980;302:1169.

7. McCarthy E, Finkel ML. Second opinion elective surgery programs: outcome status over time. Med Care 1978;16:984.

8. Paris M, Salsberg E, Berenson L. An analysis of nonconfirmation rates. JAMA 1979;242:2424.

9. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Company, 1969.

10. Campbell DT. Reforms as experiments. Am Psychol 1969;24:409.

11. Brooke RH, Williams KN, Rolph JE. Controlling the use and cost of medical services: the New Mexico Experimental Medical Care Review Organization/a four year case study. Med Care 1980;16(suppl):9.

Appendix

In this section we illustrate our interpretation of changes in surgery rates attributable to CPES with graphs of the time series for hysterectomies and tonsillectomies/adenoidectomies in Bay State, Central Massachusetts and the control area (i.e., rates for Western, Charles River and Southeastern Massachusetts combined for the period before CPES was implemented in these areas). For each time series, the following issues were considered:

> Did the fee cut (month 9) cause a statistically significant change in surgery rates? If so, pre-fee cut data are eliminated to de-

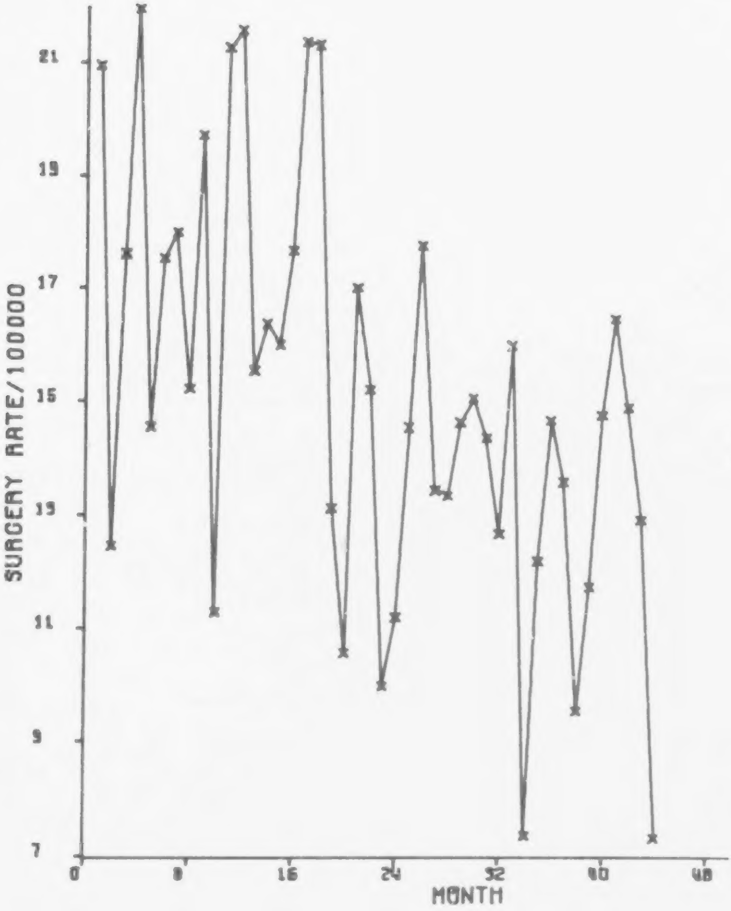


FIG. A-1. Monthly rate of hysterectomies in Bay State.

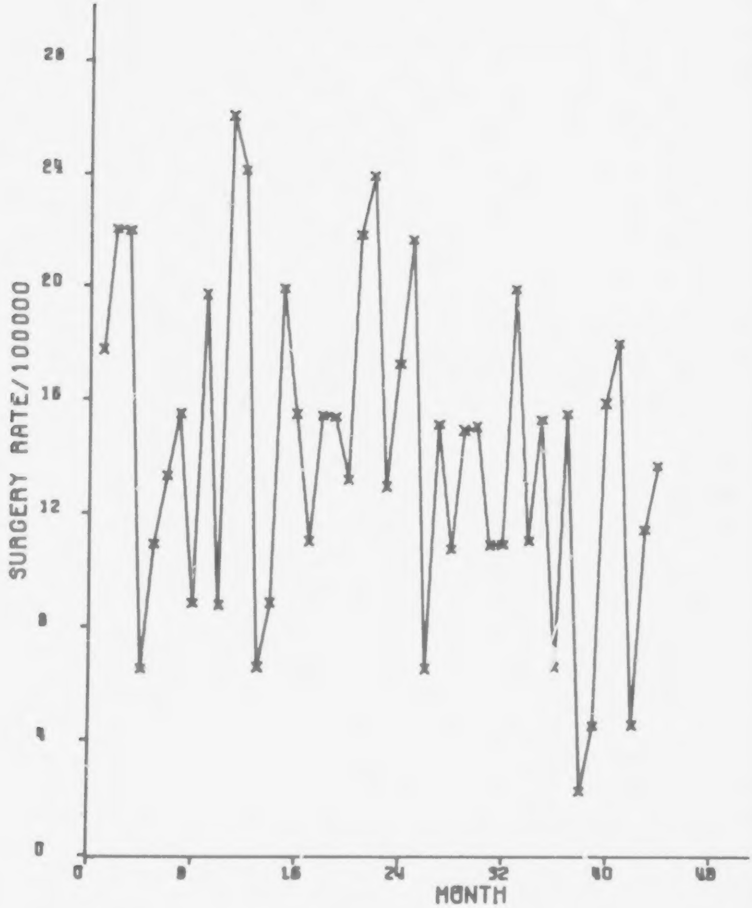


FIG. A-2. Monthly rate of hysterectomies in Central Massachusetts.

termine the pre-CPES mean level of surgery.

Was there a statistically significant decline in the mean level of surgery following CPES implementation (month 22 in Bay State, month 29 in Central Massachusetts)? The p value for the null hypothesis that the pre-CPES mean level of surgery equaled the post-CPES mean level is reported in the discussion.

Was there a statistically significant decline in the mean level of surgery in the control area following CPES implementation in Bay State or Central Massachusetts? If so, this would suggest that some factor other than CPES may have accounted for any drop in rates in Bay State or Central Massachusetts.

Should an observed drop in surgery rates after CPES implementation be attributed to the CPES or to a general decline in surgery rates over the entire time period? To evaluate this question, we explore evidence of a declining trend in pre-CPES rates in Bay State and Central Massachusetts and in the control. If there is evidence of a trend in rates prior to CPES, we considered a model to incorporate both a trend and a program impact (i.e., a linear regression model with a dummy variable).

If all evidence supports the hypothesis that CPES affected surgery rates, we include

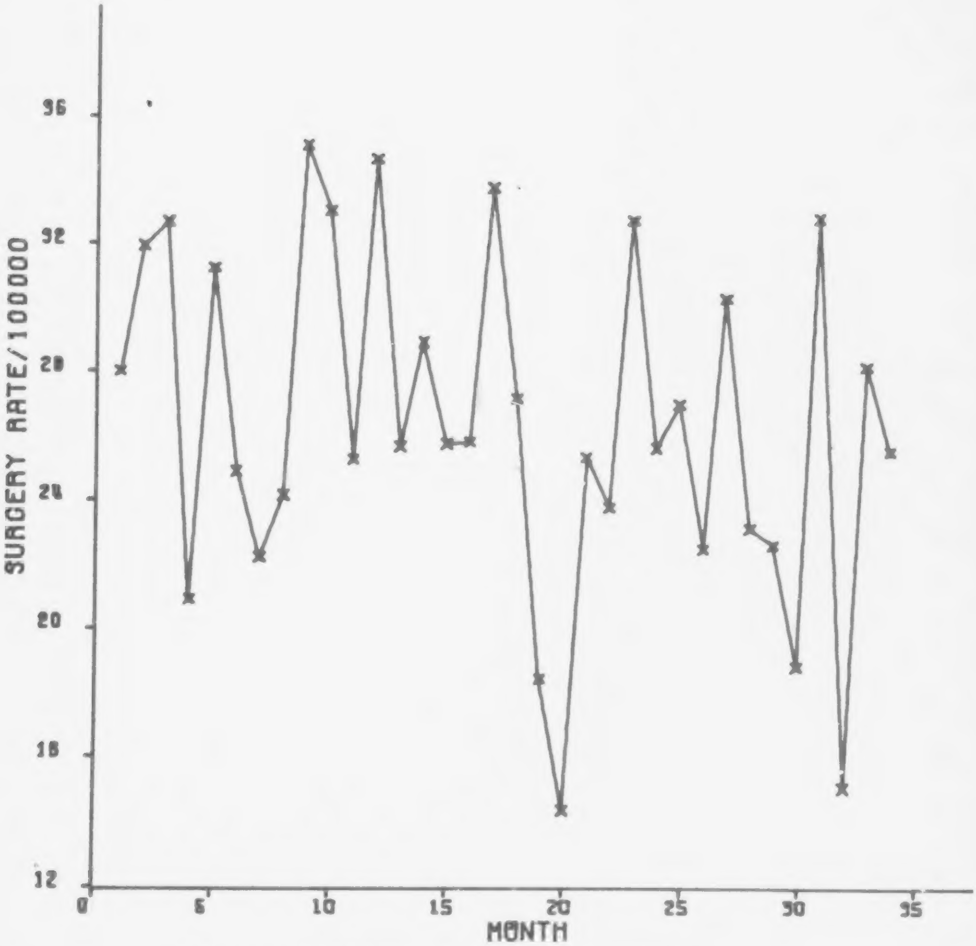


FIG. A-3. Monthly rate of hysterectomies in control area.

the effect in the minimum estimate of CPES impact. If there is a statistically significant drop in surgery rates following CPES implementation, but there is evidence to suggest that the drop may have been due to other factors, we include the effect only in the maximum estimate.

Procedures in Western, Charles River and Southeastern were considered only as candidates for inclusion in the maximum estimates of CPES impact since the post-CPES period was so short and there was no control area. The criterion for inclusion in these areas was that the change in the mean level of surgery following CPES implementaion was statistically significant.

Hysterectomies: Bay State (Fig. A-1) and Central Massachusetts (Fig. A-2)

In neither of these cases did the fee cut affect surgery rates. The mean level of surgery rates in the post-CPES period in both areas is lower than the mean level in the pre-CPES period (p = 0.00 in Bay State and p = 0.02 in Central). The drop in the

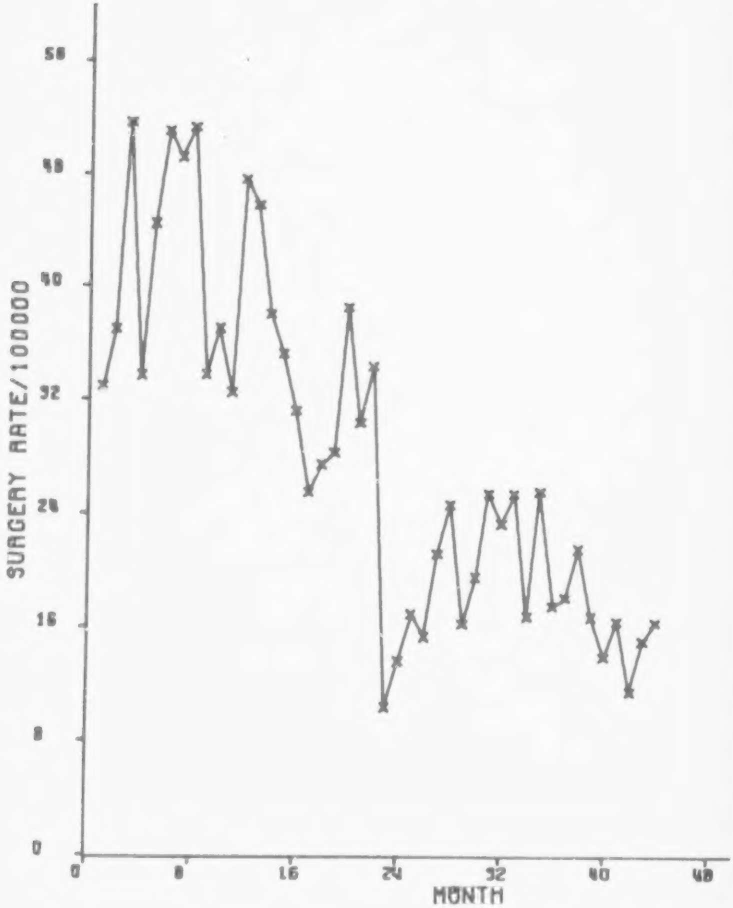


FIG. A-4. Monthly rate of tonsillectomies/adenoidectomies in Bay State.

control area following CPES implementation in Bay State and Central is not statistically significant (p = 0.20 in Bay State and p = 0.15 in Central). There is no indication of a declining trend in hysterectomy rates in the control area (Fig. A-3) or in the pre-CPES period in either region. Thus, these two cases are included in the minimum estimate of the impact of CPES.

Tonsillectomies/Adenoidectomies in Bay State (Fig. A-4)

There are several possible interpretations of this time series. The first is that the fee cut had a delayed impact (from month 9 to month 13), so that the first 13 months of data are eliminated as part of the pre-fee cut period. Surgery rates in the pre-CPES period (months 14–22) appear constant, with a significant drop in rates occurring after CPES (p = 0.00). If the first 13 months are also eliminated from the control, there is no statistically significant drop in rates there (p = 0.42), indicating that the drop in tonsillectomies/adenoidectomies in Bay State is attributable to CPES.

An alternative interpretation is that the rate of tonsillectomies/adenoidectomies

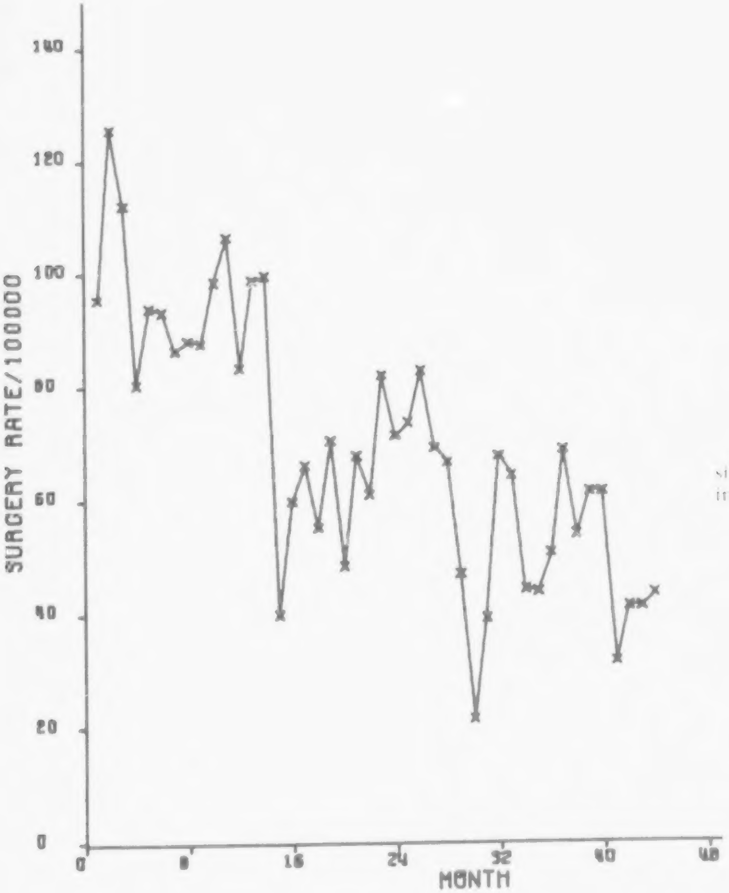


FIG. A-5. Monthly rate of tonsillectomies / adenoidectomies in Central Massachusetts.

was declining regularly over the 44-month period and that, following a temporary drop after CPES implementation, surgery rates returned to the declining trend line. This explanation implies a yearly decline of 21 per cent in surgery rates. In the control area (Fig. A-6), where the base surgery rate was over double that in Bay State, the trend line implies a yearly decline of 11 per cent. Therefore, a 21 per cent decline in Bay State tonsillectomies/adenodectomies is sufficiently unlikely that a declining trend model with no program impact seems unreasonable.

The third possibility is that the model to explain tonsillectomy/adenoidectomy rates should incorporate both a declining trend and a program impact. The yearly decline in rates is 9 per cent, in line with the 11 per cent decline in the control. The program impact predicted by this model (13.62/100,000 less surgeries per month) is almost the same as if the trend is ignored and the pre-fee cut data eliminated (12.52/100,000). Hence, whether one accepts a trend model with an impact or an impact with the trend an artifact of the fee cut and CPES, the impact of CPES is of approximately the same order of magnitude and statistically significant. Thus, this case is included in the minimum estimate.

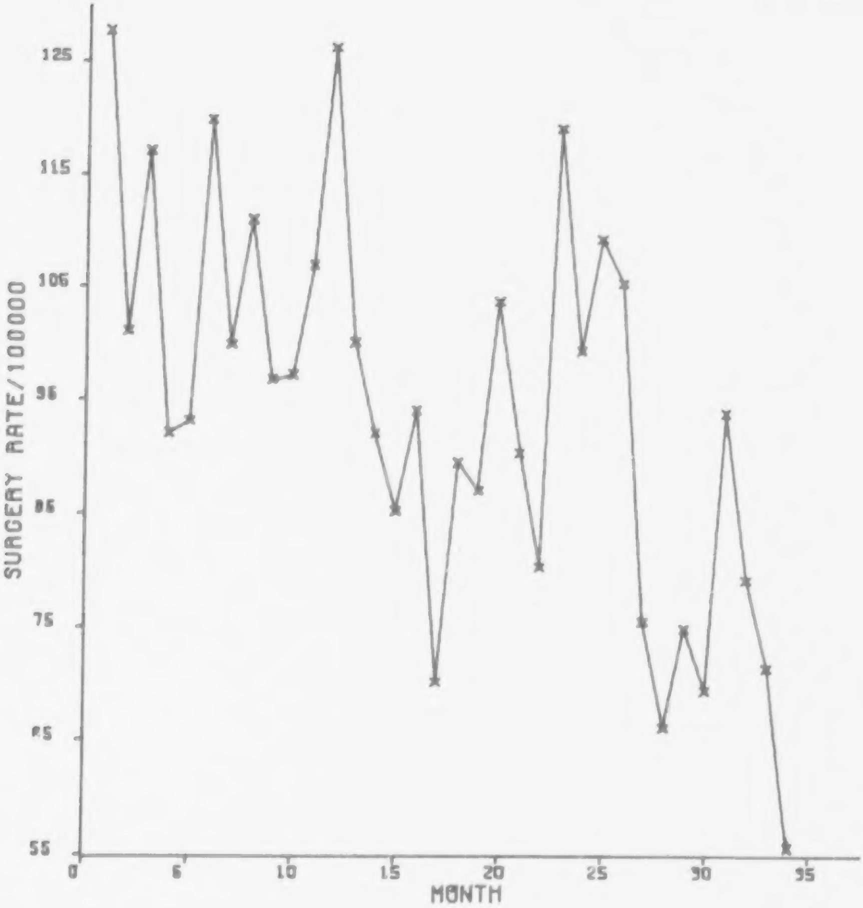


FIG. A-6. Monthly rate of tonsillectomies/adenoidectomies in control area.

Tonsillectomies/Adenoidectomies in Central Massachusetts (Fig. A-5)

If the first 14 months of data are considered part of the pre-fee cut period and eliminated, rates are constant in the pre-CPES period (months 15–29). The post-CPES mean level is lower than the pre-CPES level (p = 0.00). However, even when the pre-fee cut data are eliminated in the control area, the mean in the control area is lower after CPES implementation in Central than before CPES implementation (p = 0.03). A model that postulates a decline over the 44 months of observation implies a yearly decline of 17 per cent, more in line with the 11 per cent decline in the control, particularly given the comparability of base rates. Thus, there is evidence that other factors may have accounted for the drop in Central. Hence, this case is included only in the maximum estimate of program impact.

MEDICAL CARE
January 1982 Vol. XX, No. 1

The Effect of Concurrent Feedback in Reducing Inappropriate Hospital Utilization

JOSEPH D. RESTUCCIA, DR. P.H.

This study, conducted in four general hospitals, evaluates the effectiveness of feedback from utilization review coordinators in reducing the number of inappropriate hospital days. Experimental control was exerted over the type of feedback provided in the utilization review process to produce four treatment groups (including a control group) that vary according to 1) the channel used to inform an attending physician of a probable inappropriately located patient and 2) the amount of discretion permitted the nurse coordinator in deciding whether, when and to whom this information is provided. The results of the study indicate that physician and hospital performance in utilization review, measured both by inappropriate patient days and length of stay, are affected by the feedback strategy employed. For patients inappropriately located for a portion of their hospital stay, provision of concurrent feedback resulted in reduction of approximately two thirds of an inappropriate day and two and one half total days, compared with patients for whom no feedback was provided or for whom physician advisor involvement was required. In addition, the effect of reasons causing patients to be inappropriately located (barriers) was assessed. Barriers outside the realm of influence of the hospital or the physician were found to impede the effectiveness of the utilization review systems.

ONE OF THE EXPLICIT GOALS of the Professional Standards Review Organization (PSRO) program and an implicit goal of other cost-containment programs is to eliminate inappropriate use of health care services.[1] Achievement of this goal assumedly will have beneficial effects on both the cost and quality of health care by eliminating the expense and potential iatrogenic effects of unnecessary services. In addition, resources currently used to provide such services will be made available to fulfill unmet needs. The primary mechanism for the achievement of this goal is utilization review of the necessity for admission and continued hospital stays. Although utilization review systems have been criticized as ineffective and costly,[2-4] there has been little systematic study of the processes that affect the effectiveness of such systems.

The present study was conducted under the premises that 1) conventional review systems are largely ineffective due to their reliance on "bureaucratic" or programmed procedures that limit both the amount of communication between and the discretion of utilization review coordinators and physicians; and 2) the effectiveness of such systems is increased by reliance on "cybernetic" processes that abet communication and discretion. In particular,

---

From the Health Care Management Program, School of Management, Boston University; and the Health Care Research Section, School of Medicine, Boston University.

Supported by National Center, Grant No. 5R18-HSO 1222-02; National Center for Health Services Research, U.S. Department of Health, Education and Welfare; and W. K. Kellogg Foundation Fellowship Award.

Address for reprints: Joseph D. Restuccia, 704 Commonwealth Avenue, Boston University, School of Management, Boston, MA 02215.

0025-7079/82/0100/0046/$01.35 © J. B. Lippincott Co.

this study focuses upon the review coordinator's provision of feedback, the information about occupational performance conveyed to physicians and its relative importance in the effectiveness of hospital utilization review systems.

The study exerted experimental control over the type of information or feedback provided by coordinators to physicians in order to measure the effect of feedback on patients' inappropriate hospital days and length of stay. Before discussing the study, however, it is necessary to clarify the implications of "cybernetic" and "conventional" control mechanisms in the operation of utilization review systems.

Cybernetic Control

According to cybernetic theory, the essential means of achieving control over performance is through the use of feedback. By providing feedback, it is asserted, a system can be made to regulate itself—to exert internal or self-control. As stated by cybernetic theorist Stafford Beer[5]:

> . . . control is not a mandatory exercise, in which people are bullied or things are coerced to operate in a desired way. Rather it is a question of coaxing a system toward optimal performance; or, even better, of arranging for the system to regulate itself . . . (thus) the vital importance of the principle of feedback . . . The job of control . . . (is) less a job for the autocrat as for the steersman.

Feedback, however, is not a naturally occurring phenomenon in social systems. Communication networks must be arranged in particular patterns to insure that feedback will be provided.[6] Figure 1 illustrates how feedback can be provided in a health delivery system. The feedback loop represents information about actual system performance. This information allows comparison of actual performance and desired performance, and consequent adjustment of decisions about service delivery if a discrepancy between desired and actual performance is found. It should be noted that such discrepancies can be caused by events outside the system as well as within the system. The box labeled "uncertain events" represents exogenous variables or events in the system's environment. These events are beyond the control capacity of the system itself, yet may affect system performance in unpredictable ways. However, the feedback loop allows consequent adjustment of the system if such events cause a discrepancy between desired and actual performance. Thus, the system is structured to recognize uncertainty explicitly—to recognize the probabilistic nature of cause and effect.

The concurrent review system evaluated in this study was modeled after the cybernetic system shown in Figure 1. The basic elements of the review process of this system are contained in Figure 2. Figure 3

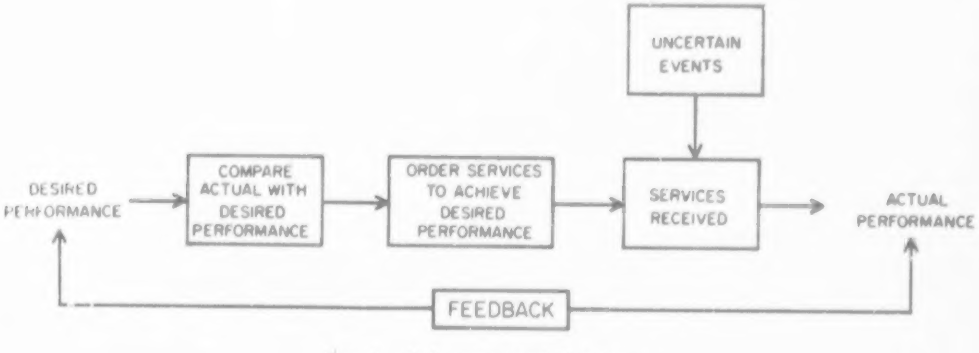

UNCERTAIN EVENTS
DESIRED PERFORMANCE
COMPARE ACTUAL WITH DESIRED PERFORMANCE
ORDER SERVICES TO ACHIEVE DESIRED PERFORMANCE
SERVICES RECEIVED
ACTUAL PERFORMANCE
FEEDBACK

Fig. 1. Cybernetic control system.

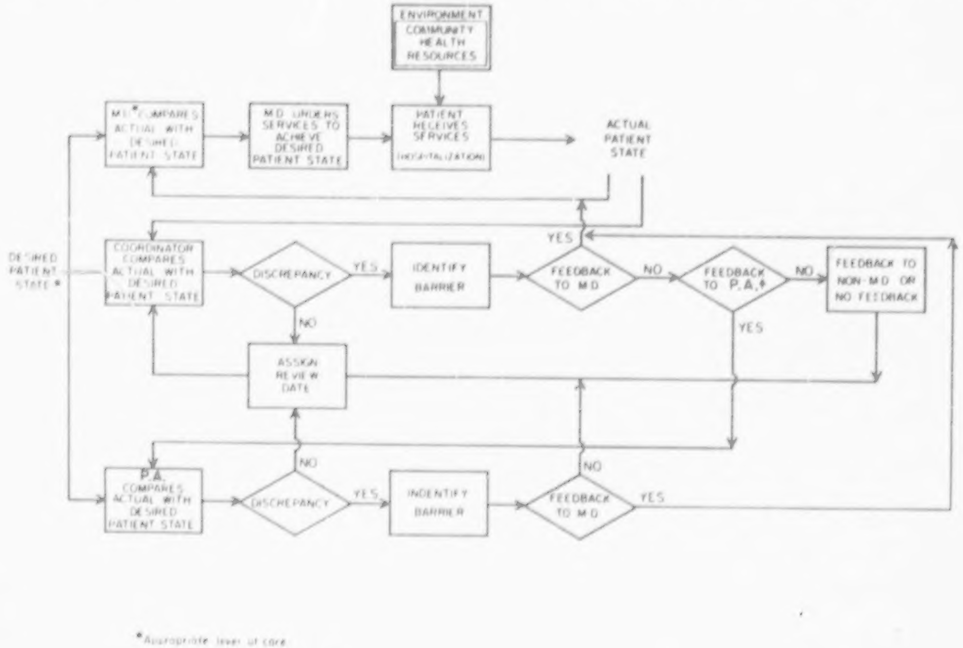
contains the basic elements of a conventional concurrent review system, modeled after federal regulations[7] and the actual procedures used in many utilization review systems.[8] A comparison of these systems reveals that each system represents alternative types of control procedures. The cybernetic system attempts to maintain control by high rates of feedback with discretion, while the conventional system attempts to control performance by programming with santions.

Comparing the "Conventional" and "Cybernetic" Utilization Review Systems

In the following section, each of the three specific hypotheses tested by this study are presented with a discussion of their relevance to the functioning of a cybernetic and a conventional concurrent review system.

Hypothesis 1: The more direct the communication channel from the utilization review coordinator to the attending physician, the more effective is the review system.

The critical element in comparing the cybernetic and conventional review systems is the difference in their communication networks. As shown in Figures 2 and 3, the most obvious difference between the two review processes is that the review coordinator provides feedback directly to the attending physician in the cybernetic system, whereas the conventional system requires that the physician contact a physician advisor if a discrepancy is found (i.e., if the comparison between the desired and actual patient state indicates that hospitalization is not necessary or if necessary hospital services are not received by the patient). If the physician advisor agrees that a discrepancy exists, the attending physician

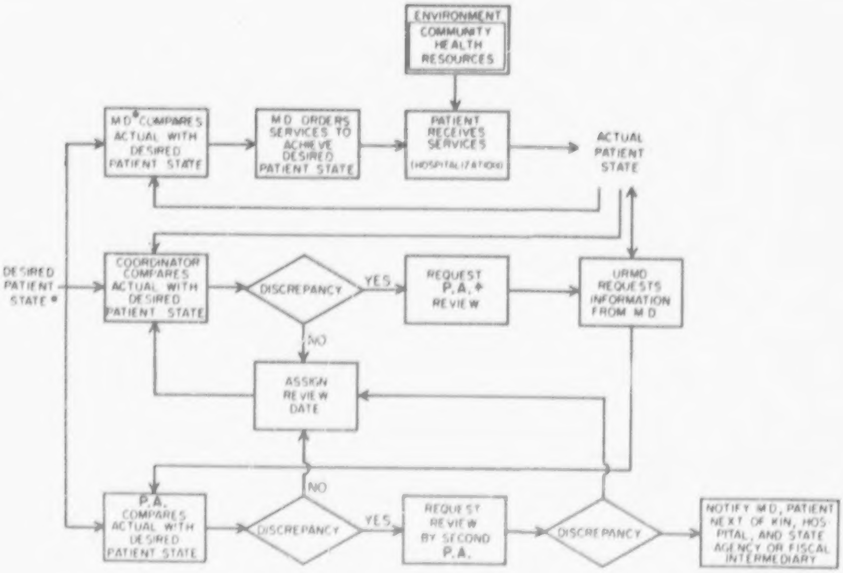

ENVIRONMENT
COMMUNITY HEALTH RESOURCES
M.D.* COMPARES ACTUAL WITH DESIRED PATIENT STATE
M.D. ORDERS SERVICES TO ACHIEVE DESIRED PATIENT STATE
PATIENT RECEIVES SERVICES (HOSPITALIZATION)
ACTUAL PATIENT STATE
DESIRED PATIENT STATE *
COORDINATOR COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
NO
IDENTIFY BARRIER
FEEDBACK TO M.D.
YES
NO
FEEDBACK TO P.A.‡
NO
YES
FEEDBACK TO NON-M.D. OR NO FEEDBACK
ASSIGN REVIEW DATE
P.A. COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
NO
YES
INDENTIFY BARRIER
FEEDBACK TO M.D.
NO
YES

*Appropriate level of care

*'M.D.' refers to attending physician

‡'P.A.' refers to physician advisor

FIG. 2. Review process in the "cybernetic" utilization review system.


ENVIRONMENT
COMMUNITY HEALTH RESOURCES
M.D.* COMPARES ACTUAL WITH DESIRED PATIENT STATE
M.D. ORDERS SERVICES TO ACHIEVE DESIRED PATIENT STATE
PATIENT RECEIVES SERVICES (HOSPITALIZATION)
ACTUAL PATIENT STATE
DESIRED PATIENT STATE*
COORDINATOR COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
REQUEST P.A.♦ REVIEW
URMD REQUESTS INFORMATION FROM M.D.
NO
ASSIGN REVIEW DATE
NO
P.A. COMPARES ACTUAL WITH DESIRED PATIENT STATE
DISCREPANCY
YES
REQUEST REVIEW BY SECOND P.A.
DISCREPANCY
NOTIFY M.D., PATIENT NEXT OF KIN, HOSPITAL, AND STATE AGENCY OR FISCAL INTERMEDIARY

*Appropriate level of care

**'M.D.' refers to attending physician

♦'P.A.' refers to physician advisor

FIG. 3. Review process in the "conventional" utilization review system.

is so informed. Thus, the feedback provided is indirect and attenuated by the likelihood that it is less timely than if provided by the review coordinator directly, due to the requirement that the physician advisor perform a review. In addition, the greater the number of people through which information is transmitted, the greater is the opportunity for the information to be blocked or distorted.[5]

Hypothesis 2: The more discretion permitted the review coordinator in deciding whether and to whom, feedback will be provided, the more effective is the system.

A second important difference is that the review coordinator in the cybernetic system attempts to identify the reason that a discrepancy exists (i.e., the "barrier" to appropriate utilization of the hospital's beds).[9] He/she determines who or what is responsible for the utilization problem—whether it is the attending physician's pattern of practice, a problem in the hospital's operating systems such as delays in scheduling tests or provision of test results, the patient's lack of cooperation with prescribed treatment or a problem in the hospital's environment such as unavailability of medical services outside the hospital. Therefore, the coordinator must obtain information about the patient's immediate environment, the hospital, and about the hospital's environment, community health care resources. The list of barriers used by the coordinators to identify reasons for inappropriate utilization is contained in Appendix I.

The significance of information about the hospital's environment is that it enables the review coordinator to determine what alternatives to a hospital level of care exist and whether the lack of such alternatives is the cause of the patient remaining in the hospital, despite the discrepancy indicating that the patient could be located elsewhere. For example, the comparison between actual and desired performances may indicate that a patient could be appropriately located in a skilled nursing facility rather than in the hospital. However, the information obtained by the coordinator to determine the cause of this discrepancy may indicate that no skilled nursing facility bed is available, thus making transfer impossible at the time of review. The cybernetic system, then, is structured to recognize uncertainty resulting from environmental events—variables outside the direct influence of the patient care system that nevertheless affect it. In contrast, the conventional system has no explicit provision for considering environmental variables until the review process has reached the stage at which the physician advisor requests further information from the attending physician, "including such factors which preclude treatment of the patient as an outpatient or in an institution providing a lesser level of care."[7]

Similarly, information obtained about the patient's immediate environment may indicate that although a patient may be inappropriately located since he/she is not receiving acute services, the patient may nevertheless require a hospital level of care. For example, the cause of the nonreceipt of services could be failure to provide ordered diagnostic or therapeutic services in a timely manner.

Thus, in the conventional system, the coordinator's decisions are programmed or restricted by a deterministic plan. He/she is little more than a monitor of performance, responsible merely for determining if a discrepancy exists and for referring discrepancies to a physician advisor for further review and resolution. In the cybernetic system, the coordinator assists in the resolution of the problem. By identifying the cause of the discrepancy, the coordinator may begin to develop a stategy to resolve it. The information about the discrepancy and this strategy can then be fed back to the attending physician. Similarly, it can be fed back to others who might assist in resolving the problem, such as discharge planners, social workers, administrators, medical consultants, nurses or personnel from relevant diagnostic or therapeutic departments. In the cybernetic system the coordinator becomes a problem solver as well as a monitor. He/she is accorded sufficient discretion to decide among multiple alternative actions: whether and when to contact the attending physician, physician advisors, other hospital departments or personnel, or any combination he/she chooses.

> Hypothesis 3: An effective system is achieved through provision of feedback by coordinators to attending physicians without the use of the sanction that is implied by the involvement of a physician advisor.

Another important difference between the cybernetic and conventional systems is that the former relies on transmission of information and the discretion of attending physicians in acting on this information, whereas the latter primarily employs negative sanctions to maintain control. In the cybernetic system, control occurs as a result of information indicating a discrepancy in system performance. The coordinator informs the attending physician that the patient *may* be "inappropriately" hospitalized, i.e., does not appear to require hospitalization or to be receiving services that would warrant continued hospitalization. Based on this information, and on the problem-solving assistance initiated by the coordinator, the attending physician presumably will take one of the following actions: 1) ensure that the patient's location be made consistent with the level of care required (i.e., discharge or transfer of the patient); 2) order the services required for the patient to be appropriately located in a hospital (i.e., order acute services available only in a hospital setting); or 3) provide further documentation justifying the appropriateness of hospitalization. Typically, the coordinator will request intervention of a physician advisor only if the attending physician fails to respond adequately to the information provided. Thus, appeal to a physician advisor is invoked only as a backup or failsafe control mechanism.

In contrast, the conventional system requires intervention of a physician advisor whenever a discrepancy is found, an intervention that is likely to be viewed as a sanction. The coordinator is invoking a "higher" authority; information to the attending physician is provided by someone of comparable occupational status—a peer rather than a subordinate. Moreover, this information is not necessarily coupled with an offer of assistance to resolve whatever problem exists. The conventional review system operates under the assumption that the attending physician alone has the responsibility and capability of resolving the problem, yet it does not offer him/her the opportunity to achieve resolution without prior evaluation and intervention by another physician. This system assumes that the physician advisor, a source of external control, must influence the attending physician to resolve the problem and that the physician advisor will be motivated to exert this influence. It fails to consider explicitly the responsibility of others for the existence and resolution of the problem or to consider the potential effect of environmental barriers in causing the problem. Little or no attempt is made to coordinate hospital and community resources that might assist attending physicians to locate patients appropriately. Thus, the conventional review system primarily relies upon programmed decisions with sanctions to achieve control, while the cybernetic system employs feedback with discretion.[10,11]

Methods

The study was conducted in four general, acute-care hospitals in the San Francisco Bay area. Table 1 provides summary characteristics of each hospital. For a two-month period in each of the four hospitals, nurse review coordinators were asked to provide different types of feedback to physicians about patients randomly assigned to one of four treatment groups. During this period, physicians, except the utilization review committee chairmen and directors of medical education, were not informed of the study to minimize reactive effects that might threaten external validity. Study patients were limited to those covered by Medicare and admitted to a medical or surgical service at one of the four hospitals. The assignment of patients was performed by the coordinators' secretaries prior to the admission review according to a systematic random sampling design.[12]

The four treatment groups represent feedback strategies that vary according to the channel used to provide feedback to physicians and the amount of discretion accorded to the coordinator in deciding whether, when and to whom feedback was provided. In each strategy the feedback decision was triggered by the coordinator's detection of a possible inappropriate patient day according to the application of the Medicare level-of-care criteria, presented in Figure 4.[13] Methodological tests of these criteria indicated that application by registered nurses trained in their use resulted in reliable (i.e., consistent) and valid judgments about appropriateness of hospital utilization.[14] Each of the six coordinators at the four study hospitals was a registered nurse who had training and experience in applying these criteria.

The four feedback strategies are as follows:

1. Direct Feedback (DF): The coordinator directly informs the attending physician that his/her patient appears to be inappropriately located.

2. Indirect Feedback (IF): The coordinator informs a physician advisor that a patient appears to be inappropriately located. It then becomes the physician advisor's responsibility to decide whether to contact the patient's attending physician.

TABLE 1. Summary Description of Study Hospitals

Hospital	Sponsorship	No. of Beds	Teaching Status	1975 Occupancy Rate (%)	Occupancy Rate During Study (%)	% of Medicare Patients	% of Medicaid Patients
1	Community	214	None	79.4	80.1	40	34
2	Religious	254	None	70.1	74.1	41	23
3	Community	437	Affiliated	78.2	74.3	36	21
4	County	584	Affiliated	57.0	56.7	21	54

3. No Feedback (NF): Neither the attending physician nor a physician advisor is informed that a patient appears to be inappropriately located.

4. Judgmental Feedback (JF): The coordinator is permitted the discretion to decide whether and in what order to inform the attending physician, physician advisor, both, or neither physician. Thus, this category may include any one or any combination of the three categories above, depending upon the judgment of the coordinator.

The four strategies were structured to represent different methods of control. Indirect feedback is closest to programming with sanctions, since it permits the coordinator no discretion and requires involvement of a physician advisor. The coordinator must follow a plan to the extent that he/she must involve a physician advisor and a sanction is invoked by the physician advisor's intervention. Judgmental feedback, in contrast, is closest to high feedback with discretion, since the coordinator's feedback decision is not programmed but discretionary, and since sanctions over the attending physician's decision are rarely invoked (assuming that the coordinator will typically choose to provide either direct or no feedback prior to or instead of providing indirect feedback).

Direct feedback is representative of an intermediary position. No sanctions are invoked, since the attending physician is informed of a patient's inappropriate location without involvement of a peer, but neither is the coordinator permitted discretion in the feedback decision. The category in which no feedback is provided serves as a control group, since no concurrent action is taken by the coordinator for inappropriately located patients in this category.

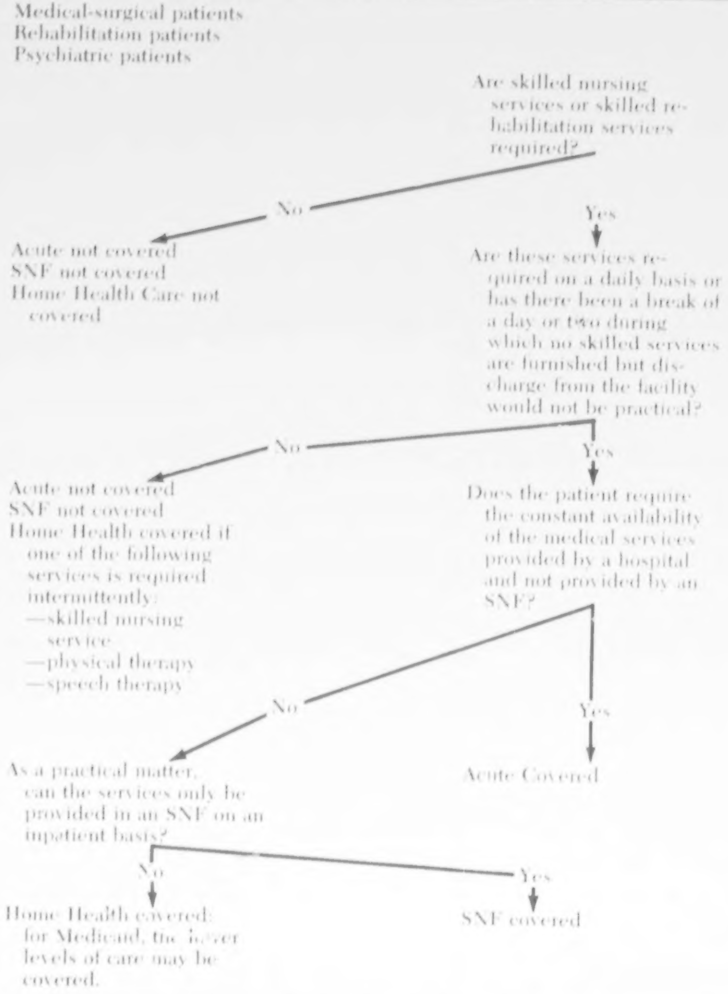
Dependent Variables

The primary dependent variable is the number of inappropriate patient days occurring for patients who experience one or more inappropriate days during their hospital stay. The variable is a direct measure of the degree of control over physician performance exerted by each "treatment" or feedback strategy in complying with the goal of eliminating inappropriate hospital utilization. The null hypothesis, then, is that there is no difference in the average number of inappropriate days in each treatment category.

A second dependent variable is the total case mix adjusted length of stay for patients in each treatment group. While a less sensitive measure of control exerted by the utilization review system than number of inappropriate days, a reduction in inappropriate days would likely result in a reduction in patients' total hospital stay.

Standardizing Patient Reviews

An important variable in utilization review is prediction of the day of the patient's stay on which review should be conducted. Studies have demonstrated that the proba-


Medical-surgical patients
Rehabilitation patients
Psychiatric patients
Are skilled nursing services or skilled rehabilitation services required?
No
Yes
Acute not covered
SNF not covered
Home Health Care not covered
Are these services required on a daily basis or has there been a break of a day or two during which no skilled services are furnished but discharge from the facility would not be practical?
No
Yes
Acute not covered
SNF not covered
Home Health covered if one of the following services is required intermittently:
—skilled nursing service
—physical therapy
—speech therapy
Does the patient require the constant availability of the medical services provided by a hospital and not provided by an SNF?
No
Yes
As a practical matter, can the services only be provided in an SNF on an inpatient basis?
Acute Covered
No
Yes
Home Health covered: for Medicaid, the lower levels of care may be covered.
SNF covered

Examples

Skilled Nursing and Skilled Rehab Services

1. Overall Management and Evaluation of Care Plan.
2. Observation and Assessment of the Patient's Changing condition.
3. Patient Education Services.
4. Direct Nursing Services
   a. IV, IM, or SC injections and IV feeding;
   b. Levin tube and gastrostomy feeding;
   c. Nasopharyngeal and tracheotomy aspiration;
   d. Insertion and new sterile irrigation and replacement of catheters;
   e. Application of dressings involving prescription medications and aseptic techniques;
   f. Treatment of extensive decubitus ulcers or other widespread skin disorder;
   g. Heat treatments;
   h. Initial phases of a regimen involving administration of medical gases;
   i. Rehabilitation nursing procedures, including the related teaching, and adaptive aspects of nursing, that are part of active treatment.
5. Direct Rehabilitation Services
   a. Therapeutic exercises or activities;
   b. Gait evaluation and training;
   c. Range of motion exercises;
   d. Maintenance therapy;
   e. Ultrasound, short-wave and microwave therapy treatments;
   f. Hot pack, hydrocollator, infrared treatment paraffin baths and whirlpool;
   g. Speech pathologist or audiologist services.

From the Federal Register, Vol. 40 (186), September 24, 1975.

FIG. 4. Medicare levels-of-care criteria.

bility of a patient day being inappropriate generally increases as length of stay increases.[15] Thus, it is necessary to "standardize" review dates to eliminate any bias (unconscious or otherwise) between treatment groups that might be introduced by the coordinators' assignment of review dates.

Standardization of review dates was achieved by employing strict length-of-stay-by-diagnosis guidelines to dictate when reviews should be made. Reviews were made in accordance with provisions of federal utilization review regulations applicable at the time of the study.[7] In general these requirements are that review be performed with 24 hours of admission and on specified percentiles (50th, 75th, 90th) of length of stay by diagnosis for a region, with a maximum of five working days between reviews irrespective of length-of-stay-by-diagnosis guidelines.

In addition, if the coordinator decided that a patient was inappropriately located, the patient was reviewed on each subsequent day of stay until his/her discharge. On each day that a review was made, the patient's level of care was determined according to the Medicare levels-of-care criteria. These consecutive reviews were necessary to measure accurately the dependent variable, the number of inappropriate days per patient. It could have been assumed that once an inappropriate day was detected, the patient would remain inappropriately located on each subsequent day of stay until discharge. However, if the patient became appropriately located in the hospital on one of these subsequent days, the dependent variable would have been erroneously inflated.*

Limitations on the Experimental Design

Two limitations were placed on the experimental design to minimize risking loss of Medicare revenue due to an increase in appropriate use as a result of the study.† First, the random assignment of patients to treatment categories was conducted so as to yield twice the proportion of patients in the Judgmental Feedback group as in any one of the other three treatment groups. Second, a limit was imposed on the length of time an inappropriately located patient was allowed to remain in the three treatment categories other than the Judgmental Feedback group. This time limit was four consecutive inappropriate days, including the first inappropriate day detected by the coordinator. Thus, if a patient in one of these three treatment categories remained inappropriate (i.e., was not discharged or had not experienced a rise in level of care that justified continued hospitalization) for four consecutive days, the coordinator was permitted to ignore the treatment category assigned and take whatever action he/she deemed necessary to prevent further inappropriate utilization. In effect, the coordinator could switch to the judgmental strategy on the fourth consecutive inappropriate day.

The first limit, the unequal sizes of the treatment groups, causes the experimental design to be nonorthogonal, or unbalanced. However, this does not cause any serious problems in analyzing results since it is possible to adjust statistically for nonorthogonal treatment effects.[16] The second limit is potentially more serious. if the mean response time of the treatment effects as measured by the dependent variable (i.e., the number of days between detection of an inappropriate day and resolution of the utilization problem) were to exceed four days, no differences would be detected among the four treatment categories; the alternative hypothesis (i.e., that significant differences exist among treatments) might be true, but might require more than four days for differences to occur. At a minimum, this four-day limit could be expected to affect results by decreasing the differences in response time that would normally occur between feedback strategies without the four-day limit. Thus, this limitation subjected the alternative hypothesis to a severe test by, in effect, decreasing the power of any statistical test to detect significant differences among feedback strategies.

---

* This "rise" from a nonacute level of care was a distinct possibility when the barrier was one related to inefficient delivery of service and the patient was assigned to a noncontrol treatment category. For example, if the barrier were a delay in a diagnostic test, the coordinator's feedback to a physician might incite corrective action (e.g., facilitate receipt of the test results) and the patient's level of care could again become appropriate. If the coordinator had not continued to review the patient, such appropriate day(s) subsequent to an inappropriate day(s) would have gone undetected. Thus, when such a rise in level of care was detected, the coordinator resumed reviews in the same manner as for a newly admitted patient.

† At the time this study was conducted, if the rate of retroactive denial of claims exceeded 5 per cent of the total Medicare patient days in any three-month period, the hospital lost its Medicare "waiver of liability" and payment for denied claims was withheld.

In addition to the limitations purposely imposed on the experimental design, other factors tended to reduce the differences among the treatment groups. These differences are likely to be reduced to the extent that anticipatory socialization of physicians occurs as a result of their awareness that the hospital has a functioning concurrent review system. Thus, their behavior is, in part, conditioned by the past history of the hospital review system regardless of feedback strategy. In addition, some control might be exerted as a result of potential, retroactive denial of payment for patients that the Medicare fiscal intermediary judged inappropriately located on the basis of claims and records review. This sanction, however, usually occurs at least two months after the claim is submitted and is an infrequent event (rarely having exceeded 2 per cent of total Medicare patient days in the study hospitals).

To the extent that differences between treatment categories are reduced by anticipatory socialization, retroactive denials and limitations imposed on the experimental design, the study hypotheses are subjected to a more difficult test. If differences in the number of inappropriate days nevertheless occur, there is strong evidence that they are attributable to the differences in the feedback strategies.

Results

In the four participating hospitals, the coordinators reviewed a total of 1,456 Medicare medical and surgical service patients during the study. Three hundred twenty-four patients (22.3 per cent) were found to have had at least one inappropriate day, with an average of 2.85 inappropriate days (21.1 per cent), in their average total length of stay of 13.49 days.As would be expected, the average length of stay of patients with inappropriate days was considerably longer than that of patients with no inappropriate days: 13.49 days (4,371 days/324 patients), compared with 9.31 days (10,540 days/1,132 patients). The average length of stay of all patients reviewed was 10.24 days, a figure approximately equal to the average length of stay of all Medicare patients in the four study hospitals during the year of the study.

Effect of Feedback Strategies

To determine the effect of feedback strategy on the dependent variable, the number of inappropriate patient days (IDAY), an analysis of variance (ANOVA) was conducted on 324 of the 333 patients who experienced at least one inappropriate day (nine patients who died during the study were excluded from the analysis). In order to focus on the effect of the feedback strategy provided, two independent variables were included in the ANOVA: the feedback strategy and the hospital. Inclusion of hospitals (with each hospital constituting one level) as a second independent variable removed variance in the dependent variable attributable to differences between the hospitals. In addition, a log transformation was performed on the dependent variable so that it more closely conformed to a normal distribution. Results are reported for both the transformed (IDAY) and untransformed (LGIDAY) dependent variable: the former for statistical inference, the latter for estimation of treatment effects.[16]

As shown in Table 2, the difference in LGIDAY among the four feedback strategies is statistically significant. IDAY in the judgmental and direct feedback strategies is approximately one half day less than in the control group. The indirect strategy, in contrast, resulted in no reduction in inappropriate days compared with the control group.‡

A two-way ANOVA, with feedback strategy and hospital as independent variables, was also performed to determine the effect of feedback on average length of stay (LOS). To adjust for case mix differences among patients, the Western United States PAS region average length of stay according to final diagnosis, age group, presence or absence of surgery, and presence or absence of multiple diagnoses was used as a covariate.[17]

In addition, a log transformation was performed on length of stay to approximate a normal distribution of this dependent variable (LGLOS). The results of this analysis parallel those that used inappropriate days as a dependent variable, with the exception that the direct feedback strategy was less influential in reducing total hospital stay than in reducing inappropriate days. As shown in Table 2, the major difference in length of stay is between the judgmental feedback strategy and both the indirect strategy and the control group.§

Effect of Feedback Strategies Excluding Patients with Environmental Barriers

To determine the extent to which environmental barriers impeded the effectiveness of feedback strategies, an ANOVA on LGIDAY and LGLOS was conducted excluding those patients who had inappropriate days associated with barriers beyond the immediate control of the physician, hospital, patient and patient's family. As shown in Table 3, the mean number of inappropriate days for the 192 patients without environmental barriers is 2.70, compared with 2.85 for all 324 patients with inappropriate days (and 3.07 for the 132 patients with environmental barriers).

The exclusion of patients with environmental barriers results in the same pattern of means found with all patients included, but with larger differences and higher significance levels between means. IDAY in the judgmental and direct strategies, respectively, is approximately 1.5 days and one day less than in the control and indirect feedback groups, while no difference was found between the control group and the indirect strategy.‖ The difference of over 3.5 days in length of stay between the judgmental and control groups shows the influence of the coordinator's discretion in reducing length of stay, particularly when not contrained by environmental barriers. The provision of direct feedback also exhibits an average length of stay in the hypothesized direction, although differences with other group means are not statistically significant.#

Comparison of Hospitals

Because of the small sample size of hospitals ($N = 4$) and the impossibility of controlling interorganizational characteristics

---

‡ Scheffe's test applied to the contrasts JF–NF, JF–IF, DF–NF and DF–IF are each significant at the $p \le 0.05$ level. In other words, the differences are primarily attributable to the lower number of inappropriate days in the judgmental and direct categories as compared with the indirect category and control group.

§ Scheffe's test applied to the contrast JF – (IF + DF)/2 is significant at the $p \le 0.10$ level.

‖ Scheffe's test indicates differences in inappropriate days excluding environmental barriers at the $p \le 0.01$ level for contrasts JF–NF, JF–IF and DF–IF and at the $p \le 0.05$ level for the contrast DF–NF.

Scheffe's test indicates length-of-stay differences at the $p \le 0.05$ and $p \le 0.01$ levels for the contrasts JF–IF and JF–NF, respectively.

Table 2. Average Number of Inappropriate Days and Length of Stay by Feedback Strategy

Strategy	N	IDAY*	LGIDAY*	LOS†	LGLOS†
Judgmental	145	2.59	0.48	12.23	1.05
Direct	68	2.75	0.48	13.81	1.10
Indirect	60	3.25	0.56	15.23	1.14
None (control)	51	3.25	0.57	14.59	1.12
Total	324	2.85	0.51	13.49	1.09
			$F = 2.88$ $P \leq 0.04$		$F = 3.10$ $P \leq 0.03$

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

that might affect inappropriate utilization in the study hospitals, conclusions drawn from a comparison of hospital differences must be regarded as tentative. However, certain consistencies with the results of the comparison by feedback strategy allow inferences to be made about the causes of the hospital differences.

Table 4 presents the average number of inappropriate days and the average length of stay by hospital. The ANOVA on the log of the values in Table 4 showed no significant differences among hospitals in the number of inappropriate days, whether for all patients ($F = 1.88$; N.S.) or for those without environmental barriers ($F = 1.11$; N.S.). Length of stay adjusted for case mix (again using regional average length of stay by diagnosis as a covariate) was found to be different for all patients ($F = 11.71$; $p \leq 0.000$) and for those without environmental barriers ($F = 6.93$; $p \leq 0.000$). The length-of-stay differences are primarily due to the shorter stay at Hospital 1. Also, despite no significant differences in inappropriate days among hospitals, Hospital 1 had fewer inappropriate days than did the other three hospitals.

The relative effectiveness demonstrated by Hospital 1 might be attributable to such factors as occupancy rate pressures, types of patients or modes of medical staff practice. Those factors are particularly likely to be relevant in comparing Hospital 1 with Hospital 4. Hospital 1 is a medium-sized, community general hospital with an occupancy rate of 80 per cent. Hospital 4 is a large, county general hospital with an occupancy rate of 57 per cent. Hospital 4, then, would have less pressure to dis-

Table 3. Average Number of Inappropriate Days and Length of Stay by Feedback Strategy Excluding Patients with Environmental Barriers

Strategy	N	IDAY*	LGIDAY*	LOS†	LGLOS†
Judgmental	84	2.07	0.43	11.13	1.02
Direct	43	2.58	0.46	13.18	1.10
Indirect	40	3.64	0.60	13.68	1.11
None (control)	25	3.51	0.57	14.89	1.13
Total	192	2.70	0.49	12.61	1.07
			$F = 5.62$ $P \leq 0.001$		$F = 3.41$ $P \leq 0.02$

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

TABLE 4. Average Number of Inappropriate Days and Length of Stay by Hospital

Hospital	All Patients with Inappropriate Days Included			Patients with Environmental Barriers Excluded		
	N	IDAY*	LOS†	N	IDAY*	LOS†
1	54	2.26	12.06	47	2.17	9.01
2	106	2.99	13.32	49	2.91	13.69
3	121	2.72	14.26	70	2.91	13.69
4	43	3.61	15.21	26	2.70	14.18
Total	324	2.85	13.49	192	2.70	12.61

* Adjusted for hospital differences.

† Adjusted for hospital and covariate differences.

charge patients quickly or resolve problems of inappropriate utilization due to a need for beds. In addition, being a county hospital, it has a patient mix with a high proportion of medical, social and/or financial problems that would make appropriately locating such patients difficult. Finally, the extensive teaching programs of Hospital 4, as well as other differences in the composition of its medical staff, may result in a less effective utilization review system than that of Hospital 1.

Without further investigation we can only speculate about whether and to what extent these factors contribute to inappropriate utilization. Moreover, the difference between Hospital 1 and Hospitals 2 and 3 are not as pronounced relative to such factors as the differences between Hospital 1 and Hospital 4. However, based on the results of the comparison of strategies, there are factors that can be examined to explain (at least in part) differences in the relative effectiveness of the utilization review systems in the four hospitals.

First, our results show that direct feedback is more effective than indirect or no feedback. Thus, if there are differences between the hospitals in the amount of direct feedback provided to the physician, we would expect this to be reflected in the measures of utilization review system performance. Second, the more discretion permitted (or assumed by) the coordinator, the more likely is the utilization review system effectiveness. Finally, the greater the proportion of environmental barriers causing inappropriate utilization, the less control the utilization review system is capable of exerting and, thus, the more likely is inappropriate tuilization.

Table 5 presents measurements of these three factors. Column 1 indicates the feedback strategy chosen by the coordinators at each hospital for patients assigned to the Judgmental Feedback group. At Hospital 1, the proportion of direct feedback provided was the highest of the four hospitals, while the proportion of times in which no feedback was provided was the lowest.

Column 2 indicates the percentage of cases in which the coordinator did not adhere to the study design, primarily by failing to provide feedback before expiration of the four-day time limit. Nonadherence was substantially higher in Hospitals 2 and 3 than in Hospitals 1 and 4. The major reason for this difference is that in Hospitals 2 and 3, the coordinators' work load exceeded their capacity to meet it. This reason also helps to account for the differences in feedback strategy chosen by the coordinators (column 1). At Hospitals 2 and 3, the coordinators' excess work load precluded their exercising discretion in providing feedback as often and as promptly as they would otherwise have chosen to do so. Further evidence of this problem is that shortly after the study, both Hospitals 2 and 3 added additional coordinators to handle the work load.

Finally, as shown in column 3 of Table 5, the proportion of environmental barriers is considerably lower in Hospital 1 than in the other three hospitals. Moreover, the proportion of physician barriers is considerably higher—precisely the type of barriers we would expect to be most amenable to resolution by providing feedback to physicians. In part, this low proportion of environmental barriers is due to more available long-term care facilities and home health services in the geographic service area of Hospital 1 than in the areas of the other study hospitals. In addition, as indicated in column 4 of Table 5, Hospital 1 had more experience providing feedback to attending physicians and reducing barriers to appropriate utilization through, for example, increased efforts in discharge planning.[9] Thus, it is not surprising that its utilization review system manifests a high degree of effectiveness.

It remains to be explained, however, why the utilization review system of Hospital 4 is not more effective. Of the four study hospitals, Hospital 4 has the highest average number of inappropriate days and the longest length of stay (Table 4). Yet, on the factors predicting utilization review system effectiveness (Table 5), it ranks second only to Hospital 1.

The reason for this apparent discrepancy was touched upon earlier when comparing Hospitals 1 and 4. As a county general hospital, Hospital 4 tends to receive the most difficult utilization problems—patients with little or no financial resources to pay for medical care and social problems that make appropriate location difficult. These problems should be reflected in the proportion of environmental barriers encountered. Yet this proportion is actually lower than and equal to those of Hospitals 2 and 3, respectively. However, when we look at multiple barriers, we find that 23 per cent

TABLE 5. Predictors of Study Hospitals' Utilization Review System Performance

Hospital	Type of Feedback* Provided for Patients in Judgmental Group (% of Patients)		Adherence to Study Design (% of Patients)		Distribution of Barriers by Area of Responsibility (% of Patients)		No. of Years Using "Cybernetic" UR System
1	Direct	65			Physician	61	3
	Indirect	12	Adherence	91	Hospital	08	
	None	23	Nonadherence	09	Patient	13	
					Environment	17	
2	Direct	37	Adherence	76	Physician	38	0.50
	Indirect	4	Nonadherence	24	Hospital	03	
	None	59			Patient	03	
					Environment	57	
3	Direct	33			Physician	42	0.25
	Indirect	2	Adherence	71	Hospital	03	
	None	65	Nonadherence	29	Patient	08	
					Environment	47	
4	Direct	50			Physician	41	2
	Indirect	20	Adherence	91	Hospital	06	
	None	30	Nonadherence	09	Patient	06	
					Environment	47	

* In cases in which the coordinator provided both direct and indirect feedback, only the type provided first was included.

of the inappropriately located patients at Hospital 4 had more than one barrier that required resolution before the patient could be appropriately located. In Hospital 1, 2 and 3, this proportion was only 5, 6 and 7 per cent, respectively.

In addition, the environmental barriers encountered by Hospital 4 were less amenable to resolution than were those at the other hospitals. This is indicated by the difference in inappropriate days found when including and excluding patients with environmental barriers: 0.91 days for Hospital 4 (3.61 − 2.70) compared with −0.09, 0.08, and −0.19 in Hospitals 1, 2 and 3, respectively (Table 4). Thus, environmental penetration—factors beyond the control of the hospital or its utilization review system—account for much of the relative ineffectiveness exhibited by the utilization review system at Hospital 4.

Summary and Discussion

For all 1,456 patients reviewed in this study, the average difference in length of stay between the Judgmental Strategy and the control group is slightly more than one half day per patient (22.3 per cent of 1,456 patients with inappropriate days × 14.59 days − 12.23 days = 0.52 days/patient). The average direct cost of the utilization review systems at the four hospitals was less than $10 per discharge. If the marginal cost savings associated with a reduction of one half day per patient in length of stay exceeds $10, the utilization review systems at these hospitals have a positive benefit–cost ratio. Using an HEW estimate of marginal cost savings per Medicare day of $87.00 for the San Francisco PSRO area in 1977,[18] the benefit–cost ratio is in excess of 4 ($87.00/2: $10.00 > 4 : 1). Furthermore, in this study no attempt was made to "focus" review on patients with a high probability of experiencing inappropriate days. Similarly, no attempt was made to select review dates within a patient's length of stay (other than percentile of length of stay by diagnosis) to maximize the probability of detecting an inappropriate day. Lower review costs can be achieved with such systems for focused review[8] and review date selection.[14]

In general, the study hypotheses, although slightly revised, have been confirmed.

Hypothesis 1: The more direct the feedback channel, the greater is the control over professional performance.

When coordinators directly informed attending physicians that a patient appeared to be inappropriately hospitalized, the physicians acted to resolve utilization problems more rapidly than when no information was given or when coordinators involved a physician advisor. As previously discussed, direct feedback conveys more timely and accurate information. In addition, as Blau and Scott[19] assert, high rates of communication tend to improve decision-making processes through 1) relief of anxiety about decisions resulting from social support; 2) provision of error correction resulting from different approaches of different people to solving a problem; and 3) incentives for advancing good suggestions and criticizing poor suggestions resulting from competition for respect in judgments. The consequent increase in ability to solve problems is particularly relevant in this situation since explicit in the coordinator's feedback is his/her assistance in identifying and resolving the utilization problem.

Hypothesis 2: The more discretion allowed in the feedback decision, the greater is the control over professional performance.

When coordinators were permitted to decide whether, when and to whom feedback was provided, the time to resolve utilization problems was shorter than when they were restricted to providing a particular type of feedback. A similar finding was made by Gertman and Egdahl[8] in a case study of utilization review activities in 44 Massachusetts hospitals.

One reason for this finding is that the coordinators preferred the responsibility and enhanced professional status associated with discretion and were thus motivated in their job performance. A second reason is that the coordinators felt confidence in the levels-of-care criteria and consequently in their appropriateness decisions made by applying these criteria. In effect, the criteria helped the coordinators to decide when discretion was warranted and helped to increase their confidence in exercising discretion. Furthermore, it is likely that limiting the coordinators' repertoire of potential responses to utilization problems frequently resulted in inappropriate responses. On one extreme, for example, the coordinator might encounter a routine case of inappropriate utilization, such as a patient kept for an additional day waiting for transfer to a skilled nursing facility. To require any response at all (other than reviewing the patient the next day to see if the transfer has occurred) would be unnecessary and, possibly, an impetus for physician resistance. On the other hand, the coordinator may encounter a case with a serious barrier to appropriate utilization, such as a delay in performing a critical test or procedure. In this situation, the attending physician—and possibly others involved in the patient's treatment—should be informed, the cause of the delay identified and corrective action taken as soon as possible.

Hypothesis 3: An effective system is achieved through provision of feedback by coordinators to attending physicians without the use of the sanction that is implied by the involvement of a physician advisor.

The relative effectiveness of direct and judgmental feedback supports this hypothesis. The relative ineffectiveness of indirect feedback indicates that involving a physician advisor caused no control to be exerted over attending physicians' performance. These results call into question the conventional concept of "peer review" which asserts that only a peer (e.g., a physician) can review and positively effect change in the behavior of other peers. This is not intended to imply that peer review is unnecessary, but rather that in level-of-care determinations a nonphysician, guided by explicit criteria and medical staff support, can positively influence physician performance. In addition, the degree of influence is greater than that exerted by a conventional or "bureaucratic" system that requires peer intervention. Thus, *routinely* involving physician advisors to corroborate the coordinator's level-of-care decision and to provide feedback to attending physicians is not only unnecessary but detrimental to utilization review system performance. Physician advisors at all four study hospitals were reluctant to confront their peers when informed of utilization problems, a finding similar to that of Sanazaro and Worth[20] in a study of concurrent quality assurance in five PSRO areas. A more effective and certainly less costly use of physician advisor time would be for the physician advisor to act as a consultant employed to advise and assist the coordinator in cases seen by the coordinator as requiring the expertise or authority of a physician.

The study, then, has demonstrated that self-regulation of professional performance can be achieved. Requisite to self-regulation is rendering visibility to performance. This visibility, in turn, requires an information system capable of comparing desired and actual performance and of providing timely feedback about discrepancies in performance. When such information was provided—along with sufficient discretion to the provider of the information, identification of the causes of discrepancies (i.e., barriers) and problem-solving assistance—physicians typically responded with action that corrected the discrepancy, without resort to sanctions. This finding is particularly relevant to hospitals, organizations having little access to sanctions over physicians, the professional group upon whom they are most dependent yet whose performance they are expected to control.

The potential amount of control that can be exerted, however, is also dependent upon the extent to which barriers are within the scope of physicians' and hospitals' realms of influence. Environmental penetration—intrusion of social, economic and financial barriers into the confines of the hospital—restricts the ability of control systems to achieve desired performance. A significant amount of control, then, can be exerted by a system operating on the basis of cybernetic principles, with the upper limit on the amount of control determined by the extent of environmental penetration.

Acknowledgments

The author is indebted to the utilization review coordinators and the medical and administrative staffs of the four study hospitals, and to Don C. Holloway, Ph.D., for his advice and encouragement.

References

1. Goran MH. The evolution of the PSRO Hospital Review System. Med Care 1978;17(Suppl):1.

2. Averill MA, McMahon LF. A cost benefit analysis of continued stay certification. Med Care 1977;16:158.

3. Dobson A, et al. PSROs: their current status & their impact to date. Inquiry 1978;15:113.

4. McSherry CK. Quality assurance: the cost of utilization review and the educational value of medical audit in a university hospital. Surgery 1976;80:122.

5. Beer S. Decision and control. London: Wiley, 1966:255.

6. Warner DM, Holloway DC. Frameworks for quantitative decision making & cybernetic control. In: Decision making and control for health administration. Ann Arbor: Health Administration Press, 1978.

7. U. S. Department of Health, Education and Welfare, Social and Rehabilitation Service. Utilization review. Federal Register March 30, 1976;41:13452.

8. Gertman PM, Egdahl RE. The dynamics of utilization review: a case study of 44 Massachusetts hospitals. Ann Surg 1978;188:544.

9. Restuccia JD, Holloway DC. Barriers to appropriate utilization of an acute facility. Med Care 1976;14:559.

10. March JG, Simon HA. Organizations. New York: Wiley, 1958.

11. Hage J. Communication and organizational control: cybernetics in health and welfare settings New York: Wiley, 1974.

12. Hess I, Reidel DC, Fitzpatrick TB. Probability sampling of hospitals & patients. 2nd ed. Ann Arbor: Health Administration Press, 1975:15.

13. Holloway DC, Holton J, Goldberg GA, Restuccia JD. Development of hospital levels of care criteria. Health Care Management Review 1976;1:61.

14. Fuhrer R. A level of care index: development & evaluation of a computer aided utilization review selector method. Unpublished dissertation, Section on Medical Information Science, Division of Ambulatory & Community Medicine, University of California, San Francisco, June 1977.

15. Gertman PM, Bucher BM. Inappropriate hospital bed days and their relationship to length of stay parameters. Unpublished paper presented at the Medical Care Section, 99th Annual Meeting of the American Public Health Association, Minneapolis, October 11, 1969.

16. Neter J, Wasserman W. Applied linear statistical models. Homewood, IL: Irwin, 1974.

17. Commission on Professional and Hospital Activities. Length of stay in PAS hospitals, United States, Western Region, 1975. Ann Arbor, 1976.

18. Health Care Financing Administration. PSROs 1978 Program Evaluation. Office of Policy, Planning, and Research, Department of Health, Education and Welfare, Pub. No. HCFA-03000, January 1979.

19. Blau PM, Scott WR. Formal organizations: a comparative approach. San Francisco: Chandler, 1962.

20. Sanazaro PJ, Worth RM. Concurrent quality assurance in hospital care: report of a study by private initiative in PSRO. N Engl J Med 1978;298:1171.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Medical Information Systems: Assessing Impact in the Areas of Hypertension, Obesity and Renal Disease

JAMES L. ROGERS, PH.D.,* OLGA M. HARING, M.D.,† PAUL M. WORTMAN, PH.D.‡
ROYCE A. WATSON, M.B.A.,§ AND JOHN P. GOETZ, B.A.†

This study describes the influence of a computerized medical record summary system in three disease areas (hypertension, obesity and renal disease) observed in the course of a controlled, randomized and prospective study of 479 Northwestern University Clinic patients. Experimental patients, who had available automated record summaries, and control patients, who had available only the manual record, were compared on several medical tests and procedures whose yearly occurrence was considered good medical practice for this patient population, and were compared as well on several measures of outcome of medical care. Evidence suggesting better care and outcome of care among patients with computerized record summaries available is presented.

THERE IS CURRENTLY GREAT INTEREST in computer-automated medical records, and technical descriptions of the various medical information systems (MIS) and their capabilities are available.[1–16] It appears that computerized records are accepted by physicians, improve access to information, facilitate patient management and research and provide educational opportunities.[5, 9, 10, 14, 17–23]

Relatively few studies examine the impact of MIS on the actual practice and outcome of patient care. Bulpitt et al.[18] found no difference between hypertensive patients with and without automated record summaries on blood pressure measurements, dropout rates and the frequency of most laboratory investigations. McDonald[20–21] found that automation leads to fewer physician errors and that these errors occur in the first place due to the physicians' limited capacity to process information rather than to deficient medical reasoning. Finally, Rogers and Haring[24] found that although no relationship exists between the use of automated record summaries and the incidence of hospitalization, the length of hospitalization per visit declines by the second year after automated records are adopted.

Clearly, the available data are insufficient to allow the formulation of general conclusions, thus underscoring the need for objective assessments of the impact of MIS on patient care over a variety of disease areas and computer systems. The

---

Address for reprints: James L. Rogers, Ph.D., Department of Psychology, Wheaton College, Wheaton, IL 60187.

The major support for this project was provided by Grant Number HS 02649 from the National Center for Health Services Research, HRA. Initial data collection and analysis were made possible by DHEW Grant Number H500674-04 and USPAS Grant Number RR 05370 (NIH).

From the Department of Psychology, Wheaton College,* the Northwestern University Medical School,*† the School of Public Health, University of Michigan,‡ and the National Center for Health Services Research.§

0025-7079/82/0100/0063/$01.10 © J. B. Lippincott Co.

purpose of this article is to describe the influence of one MIS in three disease areas (hypertension, obesity and renal disease) observed in the course of a controlled, randomized and prospective study of 479 Northwestern University Clinic patients. Experimental and control patients are compared on several medical tests and procedures whose yearly occurrence was considered essential to good medical practice for this patient population and are compared as well on several measures of outcome of medical care.

To our knowledge, two of the disease areas under study, renal disease and obesity, have not yet been the subject of previous investigations. The third area, hypertension, has been investigated (see Bulpitt et al.[18]), but the present study covers a longer time frame, involves a different MIS and incorporates a more sensitive statistical analysis.

Methods

Computer System

We have developed at the Cardiac, Pulmonary and Renal Clinics of the Northwestern University a computerized medical record system (NUCRSS) to provide physicians with concise and current information on patients' problems, to identify omissions in recording of observations and treatment recommendations, to show ordered procedures that were not carried out, to record deficiencies in medical reasoning and, most importantly, to recommend corrective actions according to selected criteria. These criteria of "good care" were established by consensus of the clinicians providing care at our university.

Study Design

A prospective, randomized study was undertaken to evaluate the effect of the record summary on the practice and outcome of care. Patients were eligible for the study if they had attended the Cardiac, Pulmonary and Renal Clinics (CPRC) for more than six months and were not pregnant or under psychiatric care. From 1,200 eligible patients, 484 were randomly selected and assigned to either an experimental or control group. Five patients withdrew from the study before it began, leaving 241 experimental and 238 control patients. In the experimental group, patients had available a computer printout of a current NUCRSS summary in addition to the traditional medical record, while the control group had available only the handwritten, traditional medical record.

Physicians participating in the study were randomly divided into three groups: 1) those who were to see only patients with automated records available; 2) those who were to see patients without automated records; and 3) those whose patient load was approximately half with and half without automated records.

Data Base

One and two years after the entry of patients into the study, blind retrospective chart reviews were done for both experimental and control patients by trained personnel using a standardized evaluation form designed by us. Briefly, the resultant data base consisted of 1) items related to the utilization of services and to the overall quality of care (e.g., number of clinic visits, provider I.D., yearly routine physical examinations, recording of signs, occurrence and reason for hospitalization, etc.); 2) more detailed information, such as the presence or absence of recommended laboratory examinations for patients who had one or more of the following tracer diseases: hypertension, organic heart disease, renal disease, pulmonary disease, obesity and diabetes; and 3) answers to a questionnaire concerning patients' views on their own health and on the care received.

Standards of care were discussed and agreed upon by members of the Department of Medicine, relying upon various criteria. In the suggestions and comments of the record summary, physicians were reminded by the computer to adhere to these prescribed standards of care. For instance, if there was no evidence that serum potassium was checked periodically in patients on diuretic treatment, this was recommended; if potassium was low, appropriate medication was suggested; if patients with abnormal cultures were not treated, the suggestion was made to do so.

A description of the evaluation structure that was utilized to reflect the adequacy of care, along with a more comprehensive discussion of the design and data that were collected, has been published elsewhere.[25]

Data Presentation and Statistics

The data are presented in three parts in the results section that follows. First, dropout rates, baseline parameters and clinic attendance are compared across conditions. Second, selected medical tests and procedures whose yearly occurrence was considered by our physicians to be essential to the ongoing care of patients with chronic hypertension, obesity and renal disease were compared across conditions. These tests and procedures, which were monitored by the computer, are the examination of renal function (i.e., blood urea nitrogen, creatinine or creatinine clearance), the examination of potassium, the funduscopic examination and the intravenous pyelogram for hypertensive patients. For patients with renal disease we have selected the examination of renal function (as above), the urine analysis and the urine culture. Finally, whether or not a diet was given or reviewed was compared across conditions for obese patients. Chi-square statistics were used to determine whether differences across conditions in the incidence of items noted here were statistically reliable.

In each of the three disease areas, hypertension, obesity and renal disease, at least one indicator of outcome of care was collected and compared across conditions. Systolic and diastolic blood pressure measurements were recorded for hypertensive patients; the number of pounds overweight was periodically collected for obese patients; and the outcomes (normal or abnormal) during year 2 of the study for the renal function examination, the urine analysis and the urine culture were collected for patients with renal disease. The analysis of variance and the analysis of covariance were used to compare the experimental and control conditions on blood pressure and weight measurements, while chi-square statistics were used to compare patients with renal disease across conditions on the various laboratory examination results.

It is possible to compare all dependent variables across conditions by examining the various tables that follow. Values of statistics and the corresponding p levels are reported in table footnotes. Throughout the text, a p value of 0.05 or less is considered by the authors to be statistically reliable.

Results

Comparisons between the experimental and control groups are presented for 1) dropout rates, baseline parameters and clinic attendance; 2) selected medical care events that were considered a priori to reflect good medical practice for this patient population; and 3) several measurements of outcome of care.

Dropout Rates, Baseline Parameters and Clinic Attendance

Table 1 presents the number and per cent of patients in each of the three disease areas under investigation, along with the number and per cent of patients in each area who died by the end of the two-year study period or were transferred to another clinic, moved or left for unknown reasons. These figures indicate that differential dropout rates across conditions due to death or other reasons were not present.

Random assignment of subjects to conditions was conducted to equate by chance the conditions on factors other than the

TABLE 1. Patient Dropout Rates for Three Disease Areas

Disease Area/ Condition	No. of Patients with Disease (%)*	No. of Patients Dead by End of Study (%)†	Transferred to Another Clinic, Moved or Left for Unknown Reasons by End of Study (%)†
Hypertension			
Experimental	197 (81.7)	28 (14.2)†	28 (14.2)
Control	185 (77.7)	33 (17.8)	23 (12.4)
Obesity‡			
Experimental	68 (28.2)	1 (1.5)	5 (7.4)
Control	79 (33.2)	7 (8.6)	10 (12.7)
Renal Disease			
Experimental	64 (26.6)	10 (15.6)	14 (21.9)
Control	45 (18.9)	10 (22.2)	4 (8.9)

Notes: When the experimental and control conditions were compared on the above dropout rates, in no instance was the difference statistically reliable ($p > 0.05$).

* Percentages in parentheses based on total number of experimental (241) or control (238) patients in the study.

† Percentages in parentheses based on total number of experimental or control patients noted in column 1.

‡ Patients whose weight exceeded 20 per cent of their ideal wieght were classified as obese.

presence or absence of the MIS. To check for *random* differences at baseline, the following baseline parameters were compared: age; sex; race; fee schedule; prior clinic attendance in months; prior medical chart weight; the prior presence or absence of a) hypertension, b) renal disease, c) obesity, d) diabetes, e) organic heart disease and f) pulmonary disease; and the total number of diseases a through f present in each patient. As would be anticipated due to the use of random assignment, most differences between conditions on these variables were small. However, in a few instances, chance differences across conditions in baseline parameters were sufficiently large to warrant consideration as possible rival explanations for any differences on the dependent variables under study that otherwise would logically be attributed to the presence or absence of the MIS.

Baseline parameters judged by us to be potentially important are presented in Table 2, which indicates that 1) there were more hypertensive males in the control group than in the experimental group; 2) diabetes as a concomitant disease among obese patients was more prevalent in the experimental than in the control group; and 3) for all three disease areas, patients in the control conditions had been attending the Cardiac, Pulmonary and Renal Clinics prior to the beginning of the study for a longer average time than had the experimental patients. The implication of these baseline discrepancies for the interpretation of differences found on the dependent variables under study (see Tables 3 and 4) will be discussed throughout the following sections.

As a final comparison, the two conditions were contrasted on clinic attendance during year 1 and year 2 of the study. There was little difference across conditions, with both groups averaging between four and six visits per year per patient.

Medical Care Events

The variables selected to reflect important medical care events for this patient population are presented for each disease area in Table 3. Distributions indicating whether these events occurred during year 1 only, year 2 only, both years or not at all are provided for both the experimental and control conditions.

Unless otherwise noted, when differences between conditions on variables in Table 3 are cited in the following text as existing, it can be assumed that they are statistically reliable ($p < 0.05$). As noted earlier, the exact value of statistics and their corresponding p levels are presented in table footnotes, a procedure adopted throughout this and the following section.

Hypertension. Examination of the renal function and of potassium occurred more frequently during both years of the study in the experimental group. Failure to perform these examinations at all during the study period was more frequently found among the control patients. When these examinations occurred only once in the two-year study period, there was some tendency for the examinations to be done during year 1 for the experimental patients but during year 2 for the controls. There was little difference in distributions across conditions for either the funduscopic examination or the intravenous pyelogram.

Obesity. Regardless of whether the data are combined over sex or independently examined by sex, failure to give or review a diet at any time during the two-year study period clearly occurred less often among experimental patients (see Table 3).

Renal Disease. There were fewer experimental patients who had not had a urine analysis performed either year, while more experimental patients had tests performed both years (see Table 3). There were generally more failures to examine the renal function in the control condition, but the overall difference was not statistically significant. Likewise, the difference across conditions in the distribution of urine cultures was not statistically reliable.

TABLE 2. Baseline Parameters With Large Differences Across the Experimental and Control Groups for Patients With Hypertension, Obesity and Renal Disease*

Disease/Baseline Variable	Experimental	Control
Hypertension	N = 197	N = 185
Sex		
Male	80 (40.6%)	55 (29.7%)
Female	117 (59.9%)	130 (70.3%)
Prior clinic attendance (mean months)	80.9 ± 8.9†	104.1 ± 12.3
Obesity	N = 68	N = 79
Incidence of diabetes as a concomitant disease	30 (44.1%)	19 (24.1%)
Prior clinic attendance (mean months)	83.8 ± 15.6	107.2 ± 19.8
Renal Disease	N = 64	N = 45
Prior clinic attendance (mean months)	78.2 ± 15.6	96.9 ± 21.3

* Thirteen baseline parameters were compared across the experimental and control conditions: age, sex, race, fee schedule, prior clinic attendance in months, prior medical chart weight, the prior presence or absence of hypertension, renal disease, obesity, diabetes, organic heart disease and pulmonary disease, and the total number of these diseases present. The above differences were large enough to warrant further consideration.

† 95 per cent confidence intervals reported with means.

Baseline Differences. Within each disease area, various distributions for the dependent variables listed in Table 3 were cross classified along the corresponding discrepant baseline parameter(s) noted in Table 2. This was done to determine whether or not the differences cited above would remain unaltered after "controlling" for discrepant baseline parameters. In no instance was there a substantive change in the findings already reported for the dependent variables. Because no additional insight would be provided by displaying the multidimensional cross tabulations resulting from the incorporation of baseline parameters into the analysis, these figures have been omitted.

TABLE 3. Dependent Variables Reflecting Medical Care Events in Three Disease Areas for Experimental and Control Patients

Disease Area/ Procedure or Test	Done Year 1*		Done Year 2*		Done Both Years*		Not Done*	
	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control
Hypertension								
Renal function examination†	44 (22.3)	38 (20.5)	18 (09.1)	26 (14.1)	120 (60.9)	93 (50.3)	15 (07.6)	28 (15.1)
Potassium	46 (23.3)	38 (20.5)	20 (10.2)	24 (13.0)	119 (60.4)	97 (52.5)	12 (06.1)	26 (14.1)
Funduscopic	9 (09.5)	6 (03.2)	119 (59.8)	100 (52.6)	14 (07.0)	9 (04.7)	55 (27.9)	70 (37.8)
Intravenous pyelogram	13 (06.5)	13 (06.8)	45 (22.6)	60 (31.6)	78 (39.2)	59 (31.1)	61 (31.0)	53 (28.6)
Obesity								
Number of diets given or reviewed combined over sex	11 (16.2)	9 (11.4)	20 (29.4)	16 (20.3)	23 (33.8)	16 (20.3)	14 (20.6)	38 (48.1)
Males	3 (15.0)	1 (06.7)	9 (45.0)	1 (06.7)	6 (30.0)	7 (46.7)	2 (10.0)	6 (40.0)
Females	8 (16.7)	8 (12.5)	11 (22.9)	15 (23.4)	17 (35.4)	9 (14.1)	12 (25.0)	32 (50.0)
Renal disease								
Renal function examination†	12 (18.8)	6 (13.3)	2 (03.1)	6 (13.3)	45 (70.3)	25 (55.6)	5 (07.8)	8 (17.8)
Urine Analysis	21 (32.8)	9 (20.0)	7 (10.9)	9 (20.0)	30 (46.9)	14 (31.1)	6 (09.4)	13 (28.9)
Urine Culture	31 (48.4)	27 (60.0)	6 (09.4)	5 (11.1)	16 (25.0)	9 (20.0)	11 (17.2)	4 (08.9)

Notes: Statistically significant differences across the experimental and control conditions were found for the examination of renal function (χ^2 [3] = 8.9, $p = 0.03$) and potassium (χ^2 [3] = 8.2, $p = 0.042$) among hypertensive patients; number of diets given or renewed among obese females (χ^2 [3] = 10.1, $p = 0.018$) and all obese patients combined over sex (χ^2 [3] = 12.1, $p = 0.007$); and for urine cultures (χ^2 [3] = 10.5, $p = 0.015$) among patients with renal disease. All other comparisons resulted in differences that were not statistically reliable ($p > 0.05$).

* Percentages in parentheses.

† Blood urea nitrogen, creatinine or creatinine clearance.

TABLE 4. Mean Blood Pressure for Hypertensive Experimental and Control Patients

	Unadjusted Means		Adjusted Means*	
	Experimental	Control	Experimental	Control
Systolic Blood Pressure				
At beginning of study				
Males	147.7 ± 4.9† (73)‡	149.1 ± 6.4 (53)	—	—
Females	149.9 ± 4.2 (110)	153.3 ± 4.8 (123)	—	—
After 10–15 months				
Males	145.5 ± 5.6 (65)	152.7 ± 6.7 (44)	147.7	151.8
Females	148.1 ± 4.8 (103)	147.9 ± 4.5 (117)	148.6	146.5
After 22–24 months				
Males	145.0 ± 8.4 (39)	148.2 ± 9.0 (25)	144.5	146.8
Females	146.2 ± 5.5 (80)	147.0 ± 5.4 (85)	146.9	147.0
Diastolic Blood Pressure				
At beginning of study				
Males	94.9 ± 3.2 (73)	93.4 ± 4.0 (53)	—	—
Females	93.0 ± 2.4 (109)§	91.9 ± 2.6 (122)§	—	—
After 10–15 months				
Males	92.4 ± 3.2 (65)	91.7 ± 4.4 (44)	91.5	91.4
Females	91.9 ± 2.8 (102)	90.4 ± 2.6 (117)	91.7	91.3
After 22–24 months				
Males	91.6 ± 3.9 (39)	94.4 ± 7.0 (25)	90.1	94.0
Females	91.2 ± 3.1 (80)	89.6 ± 2.8 (85)	91.3	90.1

Notes: For both systolic and diastolic blood pressure, the main effects of the treatment and of sex, as well as the treatment by sex interactions, were not statistically significant for either the unadjusted means using the analysis of variance or the adjusted means using the analysis of covariance.

* Covariance adjustment for blood pressure at the beginning of the study, prior time in cardiac-pulmonary-renal clinics and patient age.

† 95 per cent confidence intervals reported with unadjusted means.

‡ Parentheses contain numbers of patients, which differ over time periods because not all patients visited the clinic during the 10–15 month and 22–24 month periods.

§ In two instances diastolic blood pressure was not recorded, even though systolic blood pressure was available.

Patient Outcome

Hypertension. Table 4 contains the average systolic and diastolic blood pressures (measured at the beginning of the study) for patients with hypertension. Also included are averages that have been covariance adjusted for blood pressure value at the beginning of the study as well as for the two discrepant baseline parameters noted earlier, length of time spent in the Cardiac, Pulmonary and Renal clinics before the onset of the study and patient age. Both unadjusted and adjusted mean values have been computed for the time periods of 10–15 months and 22–24 months after the study had begun. Unadjusted mean values present at the beginning of the study are also given in Table 4. Use of the MIS, as evidenced by the lack of variability among means in Table 4, had no effect on systolic or diastolic blood pressure for either males or females at any point in time during the study. Details of these comparisons are found in Table 4.

Obesity. Each patient's ideal weight was subtracted from his or her weight at the beginning of the study, during the 10–15 month period and during the 22–24 month period. These differences in scores reflect "pounds overweight." In comparing the experimental and control groups, it was then necessary to adjust pounds overweight taken after the study began by the amount that the patient was overweight at the beginning of the study as well as by the patient's body size (i.e., ideal weight). The analysis of covariance was used to adjust for these two factors and for the discrepant baseline parameters noted earlier—prior attendance at the Cardiac, Pulmonary and Renal clinics and the incidence of diabetes as a concomitant disease. Additionally, the total number of other concomitant diseases each obese patient had was entered as a covariate (see footnote, Table 2). The mean unadjusted and the mean adjusted pounds overweight for males and females during the 10–15 month period and during the 22–24 month period after the study began are presented in Table 5. Also, the means for unadjusted pounds overweight at the beginning of the study are presented for the various groups.

Examination of Table 5 reveals that females are more overweight than males in both the experimental and control conditions. More importantly, by 22–24 months after the study began, females in the experimental group were less overweight than female controls and males in the experimental group were considerably less overweight than male controls.

Renal Disease. As measures of outcome of care for patients with renal disease, Table 6 contains the results of renal function examinations (i.e., blood urea nitrogen, creatinine or creatinine clearance), urine analyses and urine cultures that were given during year 2 of the study. More normal test outcomes were present among experimental patients for all three of these procedures, although the difference between conditions was statistically reliable only for the urine analyses and the urine cultures.

Newly Classified Patients. Over the course of the two-year study period, 14 (5.8 per cent) of the 241 experimental patients and 10 (4.2 per cent) of the 238 control patients were newly classified as hypertensive, a difference not statistically significant ($Z = 0.6$, $p > 0.05$). Corresponding figures for newly classified obese patients are 13 (5.4 per cent) experimental and 20 (8.4 per cent) control patients ($Z = 1.1$, $p > 0.05$). The findings previously reported for hypertensive and obese patients held for newly classified patients as well. That is, there was little difference between experimental and control patients on blood pressure measurements, while the 10–15 month period experimental patients tended to be less overweight than controls (28.2 versus 36.7 pounds for males, and 22.1 versus 42.6 pounds for females, $F[1, 29] = 4.8$, $p < 0.04$). There were no experimental or control patients that were newly classified as having renal disease during the study period.

Discussion

Generally, the experimental patients underwent more laboratory procedures and/or were given (or had reviewed) more diets than the control group. These events were considered a priori as essential to good, routine care for this population of patients. Of eight procedures examined among patients in three disease areas, six were performed more frequently among experimental than among control patients (see Table 3). Of these six differences, four were statistically significant. Further, obese patients in the experimental groups

TABLE 5. Mean Pounds Overweight for Obese Experimental and Control Patients

	Unadjusted Means		Adjusted Means[*]	
	Experimental	Control	Experimental	Control
At beginning of study				
Males	50.3 ± 6.1[†] (15)[‡]	58.2 ± 13.4 (11)	—	—
Females	54.5 ± 8.3 (40)	55.0 ± 6.2 (48)	—	—
After 10–15 months				
Males	44.8 ± 8.4 (15)	56.5 ± 16.4 (11)	45.6	48.6
Females	51.5 ± 9.1 (40)	54.5 ± 7.3 (46)	52.3	55.3
After 22–24 months				
Males	34.8 ± 8.6 (11)	59.5 ± 19.0 (9)	39.3	52.2
Females	52.0 ± 10.7 (35)	54.8 ± 7.9 (35)	51.5	55.8

Notes: Analysis of covariance shows that during the 10–15 month period, neither the treatment by sex interaction nor the main effect of the treatment was statistically significant (F [1,100] = 0.55, p = 0.45, and F [1,103] = 2.4, p = 0.12, respectively) but that the main effect for sex was significant (F [1,103] = 4.1, p = 0.045). For the 22–24 month period, the treatment by sex interaction and the main effect of sex were not statistically significant (F [1,81] = 1.8, p = 0.18, and F [1,81] = 3.2, p = 0.08, respectively), while the main effect of the treatment was significant (F [1,81] = 5.4, p = 0.023).

[*] Covariance adjustment for pounds overweight at the beginning of the study, ideal weight, prior time in cardiac-pulmonary-renal clinics, presence or absence of diabetes as a concomitant disease and the total number of other concomitant diseases.

[†] 95 per cent confidence intervals reported with unadjusted means.

[‡] Parentheses contain numbers of patients, which differ over time periods because not all patients visited the clinic during both the 10–15 month and 22–24 month periods.

were less overweight, while experimental patients with renal disease had better results on year 2 urine analyses and urine cultures.

The traditional record used in the clinics contained on the average 1½ pounds of notes, laboratory test outcomes, diagnoses and other information entered over long periods of time by different physicians. This state of affairs no doubt jeopardized continuity of care, especially since patients were not necessarily seen by the same physician from visit to visit. The present MIS, in addition to condensing the medical information to items that were current, legible and relevant, provided physicians with warnings and reminders concerning good medical practices. For this reason it was expected that the MIS would have a positive effect on medical care, and it was the sole purpose of the present experiment to determine whether this occurred by examining differences between conditions on selected dependent variables. Not knowing whether any differences would be found (or what the nature of the differences would be), this study was not designed to examine the specific mechanisms underlying any effects that the MIS might have. However, in order to attribute any effects that might be found to the presence or absence of the MIS rather than to extraneous differences between conditions on uncontrolled variables, we randomly assigned patients to conditions, randomly assigned physicians to conditions and analyzed the dependent variables controlling for the few chance baseline differences that appeared large enough to present a potential threat to the internal validity of this experiment. Further, the presence of differential dropout rates and differential clinic attendance during the study was ruled out by comparing these variables across conditions. In light of these procedures and facts, the differences between the experimental and control groups on the dependent variables cited here seem reliable, and these differences suggest that the MIS has had a positive effect on patient care.

As noted, questions remain concerning the mechanisms by which changes occur, and clarifying studies in this area will be needed. To this end, our impressions of the underlying causes of some of the differences reported here may be helpful to other investigators or to those considering the adoption of an MIS. The vast differences that exist in types of systems, clinical environments and needs, patient populations and even criteria of good care must of course be kept in mind.

Results in the area of obesity management seem the clearest. Experimental patients either had new diets assigned or old diets reviewed more frequently than control patients. Correspondingly, experimental patients evidenced greater weight loss on the average than did control patients. The NUCRSS keeps track of weight loss progress and reminds physicians to review or change diets. We hypothesize that this resulted in greater awareness on the part of physicians and patients, which in turn promoted weight loss.

A secondary finding of some interest is that more progress in weight loss was made by males than females, and that even though experimental patients were less obese than control patients, after 22–24 months they remain considerably overweight (39.3 average adjusted pounds for males and 51.3 average adjusted pounds for females). This is not surprising in that most weight loss programs reported in the literature, regardless of orientation, have been only marginally successful.[26]

For experimental patients with renal disease, it is possible that because more urine analyses were ordered during the first year of the study for them than for the control group, problems indicated by these tests could have been corrected by the second year. Additionally, the increased awareness of test results that would be expected when the MIS was used should have the overall effect of inducing appro-

TABLE 6. Results of Three Laboratory Examinations for Patients with Renal Disease

Test/Result of Test	Experimental (N = 54)		Control (N = 33)	
	Done	Not Done	Done	Not Done
Renal function examination[*]	47 (87.0)[†]	7 (13.0)	31 (93.9)	2 (06.1)
Normal result	17 (36.2)[‡]		7 (22.6)	
Abnormal result	30 (63.8)		24 (77.4)	
Urine analysis	37 (68.5)	17 (31.5)	23 (69.7)	10 (30.3)
Normal result	17 (45.9)		1 (04.3)	
Abnormal result	20 (54.1)		22 (95.7)	
Urine culture	22 (40.7)	32 (59.3)	14 (42.4)	19 (57.6)
Normal result	15 (68.2)		5 (35.7)	
Abnormal result	7 (31.8)		9 (64.3)	

Notes: When the percentages of normal results were compared across the experimental and control conditions, the following one-tailed Z statistics emerge: for the Urine Analysis, Z = 3.40, p = 0.0003; for the Urine Culture, Z = 1.91, p = 0.028; and for the Renal Function Examination, Z = 1.27, p = 0.10.

[*] Blood urea nitrogen, creatinine or creatinine clearance.

[†] Percentages in parentheses based on number of patients in the study at the end of year 1 (i.e., N = 54 or N = 33).

[‡] Percentages in parentheses for outcomes based on total number of tests done during year 2.

priate medical care and thus improving results on the urine analyses and urine cultures the following year. Unfortunately, first-year results of laboratory investigations were not available to us, and this hypothesis cannot therefore be fully tested. It is not unreasonable from a medical point of view, however, to believe that abnormalities detected in the urine analysis or urine culture could often be cleared up by the second year once they were detected. On the other hand, abnormalities found in the renal examination usually would be much more difficult to eliminate. Correspondingly, a large difference between conditions in results would not be expected and, in fact, was not found for the outcomes of renal function examinations.

The data do not suggest a simple causal interpretation in the area of hypertension. In the present study, the evaluation of renal function and the examination of potassium occurred more frequently in the experimental than in the control groups. In the Bulpitt et al. study cited earlier, the incidence of the examination of potassium was reviewed, but, unlike the present study, no statistically reliable difference was found. These investigators also examined the incidence of blood glucose examinations and found a higher incidence of this procedure for experimental patients. Although the above differences suggest that important medical events for the ongoing care of hypertensive patients may occur more routinely when an MIS is available, apparently such changes have not resulted in improved outcome of care as indexed by blood pressure values. No difference on blood pressure was found by Bulpitt et al. or by us. We had expected that better information, through such means as heightening physician awareness of medication regimens or facilitating patient compliance due to better physician–patient communication, would result in reduced blood pressure. Whatever the reasons for continued high blood pressure in both groups (e.g., noncompliance of patients with medication regimens, only partially effective medications, etc.), the computer summary has not improved (or worsened) the situation.

In this study we used simple randomization (rather than stratifying the sample before randomization), followed by the statistical control of the occasional baseline differences that emerged by chance. If the experiment could be redone, it would be best to stratify the sample along the three baseline dimensions found to differ (sex of diabetic patients, presence of diabetes for obese patients and prior clinic attendance for all patients), but there was, of course, no way of knowing before the random assignment of patients that these three variables, rather than some of the other 10 baseline variables that were also retrospectively checked, would be different across conditions. Probably the best procedure for future MIS evaluations would be to select a few important baseline variables for stratification purposes to assure equality across conditions on at least these variables, and then to adjust statistically for other baseline variables as necessary, after retrospectively examining them for differences between conditions.

As a concluding comment, we note that the simmering debate presently underway over process versus outcome measures of health care quality has some bearing on the present report.[8] For example, what is often considered "process" of care (the occurrence of important medical tests or diets) was improved among hypertensive and obese patients who had the MIS available. Yet, outcome of care was not improved among hypertensive patients (i.e., blood pressure remained high) and only relatively improved among obese patients (i.e., relative to the control groups, experimental patients were statistically less overweight, but they were still considerably overweight). Our position is that as medical knowledge advances, the relationship between process and outcome of care will improve. Therefore, the worth of broadly based technologies like MIS should be judged on the basis of how well they aid the occurrence of what is currently and locally considered good medical practice (i.e., good process of care), even if current practice is less effective in terms of outcome than would be hoped. Accordingly, the present system has made a positive impact on the occurrence of medical events considered important for the present patient population and should help to insure the occurrence of any new, more efffective procedures that might develop in the future.

References

1. Barnett GO. Computer-stored ambulatory record (COSTAR). Washington D.C., DHEW Pub. No. (HRA) 76-3145, 1976.

2. Beilin LJ, Bulpitt CJ, et al. Computer based hypertension clinical records: a co-operative study. Br Med J 1974;2:212.

3. Bradshaw-Smith JH. A computer record-keeping system for general practice. Br Med J 1976;1:1395.

4. Cobelli C, Salvan A. A medical record and a computer program for diagnosis of thyroid disease. Methods Inf Med 1975;14:126.

5. Grossman JH, Barnett GO, et al. An automated medical record system. JAMA 1973;124:1616.

6. Jerome HH, Barnett GO, et al. An automated medical record system. JAMA 1973;224:1616.

7. Kennedy F, Cleary JJ, et al. "Switch": a system producing a full hospital case history on computer. Lancet 1968;2:1230.

8. Levy RP, Cammarn MR, Smith MJ. Computer handling of Ambulatory clinical records. JAMA 1964;190:1033.

9. McDonald CJ, Murray R, et al. A computer-based record and clinical monitoring system for ambulatory care. Am J Public Health 1977;67:240.

10. Mellner C, Selander H, Wolodarski J. The computerized problem-oriented medical record at Karolinska Hospital—format and function, user's acceptance and patient attitude to questionnaire. Methods Inf Med 1976;15:11.

11. NCHSR Research Digest Series: Automation of the problem-oriented medical record. Washington, D.C., DHEW Pub. No. (HRA) 77-3177, 1977.

12. Schenthal JE, Sweeney JW, Nettleton WJ. Clinical applications of large scale electronic data processing apparatus I. JAMA 1960;173:6.

13. Schenthal JE, Sweeney JW, Nettleton WJ, Yoder RD. Clinical applications of large scale electronic data processing apparatus III. JAMA 1963;186:101.

14. Schmidt EC, Schall DW, Morrison CC. Computerized problem-oriented medical record for ambulatory practice. Med Care 1974;12:316.

15. Slack WV, Hicks GP, Reed CE, Van Cura LJ. A computer based medical history system. N Engl J Med 1966;274:194.

16. Wickham JEA, Charlton CAC, et al. A computer-based record and organization system for a department of urology. Br J Urol 1975;47:345.

17. Battele Columbus Laboratories. Evaluation of medical information system in a community hospital. Washington, D.C., DHEW Pub. No (HRA) 76-3144, 1976.

18. Bulpitt CJ, Beilin LJ, et al. Randomised controlled trial of computer-held medical records in hypertensive patients. Br Med J 1976;1:677.

19. Ehrenhaft PM, Corcoran S, Harwood EA, West H. Policy implications of medical information systems. Washington, D.C.: Office of Technology Assessment, 1977.

20. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. N Engl Med 1976;295:1351.

21. McDonald CJ. Use of a computer to detect and respond to clinical events: its effect on clinical behavior. Ann Intern Med 1976;84:162.

22. Starfield B, Simborg D, Johns C, Horn S. Coordination of care and its relationship to continuity and medical records. Med Care 1977;15:929.

23. Zielstorff RD, Roglieri JL, et al. Experience with a computer-based medical record for nurse practitioners in ambulatory care. Comput Biomed Res 1977;10:61.

24. Rogers JL, Haring OM. The impact of a computerized medical record summary system on incidence and length of hospitalization. Med Care 1979;17:618.

25. Lennie FS, Ametrano DA, Haring OM. Evaluation of an automated record summary for ambulatory care. Medinfo 1974;1:115.

26. Janda LH, Rimm DC. Covert sensitization in the treatment of obesity. J. Abnorm Psychol 1972;-80:37.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Assessing Medical Practices Through PSRO Cooperative Studies

An Evaluation of Cesarean Births in Nine PSRO Areas

WILLIAM F. JESSEE, M.D., C. W. NICKERSON, M.D., AND
WILLIAM S. GRANT, for the PSRO Coordinating Group*

Nine Professional Standards Review Organizations jointly conducted an areawide medical care evaluation study of Cesarean births. Of 5,002 Cesarean births studied in 158 hospitals, only 1.1 per cent were judged to be unjustified procedures by local physician peer review. Although there was substantial geographic variation in rates of Cesarean birth (11.0%–18.3 per cent), no statistically significant pattern of variation was identified. Maternal outcomes and infant outcomes were generally good. This study demonstrates the potential for use of PSRO data in studying major issues in health care, as well as the practical constraints of such studies.

THE PROFESSIONAL STANDARDS REVIEW Organization (PSRO) program was originally developed by the federal government to assure that health care services funded under the Social Security Act are medically necessary, appropriate and of acceptable quality.[1] In the course of carrying out these statutory duties, PSROs collect a large variety of health care data that are of potential value for increasing understanding of the forces that affect health services utilization and health care quality. Despite the attention that has been devoted to the educational role of PSROs in improving health care practice,[2] their potential as resources for evaluating contemporary community medical practices has not been adequately explored. This study represents a first attempt by a group of PSROs to pool data on an issue of substantial controversy in medical practice in an effort to define further its proper role.

Considerable attention has been focused during recent years on the rising rate of Cesarean birth in the United States.[3–6] Critics of the rising rates have pointed to the higher costs, associated morbidity and potential for late adverse effects for mother

---

From the Department of Health Administration, University of North Carolina School of Public Health, and the Colonial Virginia Foundation for Medical Care.

* The participating PSROs and their Coordinating Group representatives were: Bronx Medical Services (NY) (Seymour Cohen, M.D., Veronica M. Whalen); Colonial Virginia Foundation for Medical Care (C. W. Nickerson, M.D., William S. Grant, Sandra Latham); Essex Physicians Review Organization (NJ) (James T. Carroll, M.D., Jane Majcher); Iowa Foundation for Medical Care (Charles Driscoll, M.D., Jennifer Cofer); New Mexico PSRO (Phillip Kamps, M.D., John B. Cappes, M.D., Jim Reese); Northern Illinois PSRO (Bruce E. Greenfield, M.D., John Leonard, M.D., Donna Furman); Physicians Peer Review Organization (OH) (Lawrence H. Malm, M.D., Maria Hewett); Region Six Peer Review Organization (OH) (Robert H. Baxter, M.D., Faith A. Lower); Santa Clara Valley PSRO (CA) (William F. Jessee, M.D., Emily Wieczorek).

Address reprint requests to: William F. Jessee, M.D., University of North Carolina, Department of Health Administration, 266 Rosenau Building 201/H, Chapel Hill, NC 27514.

0025-7079/82/0100/0075/$01.00 © J. B. Lippincott Co.

TABLE 1. Screening Criteria for Evaluation of Primary Caesarean Sections

Criterion Number	Elements	Std.	Exceptions	Definitions and Retrieval Sources
1	*Justification for Surgery* One or more of the following: A. Previous uterine or vaginal surgery	100%	None	*Uterine surgery*—myomectomy, metroplasty, hysterotomy, Strassman reunification procedure, repair of uterine perforation due to trauma or repair of incompetent cervix by Shirodkar or Lash procedure. *Vaginal Surgery*—anterior vaginal repair, repair of cystocele, posterior vaginal repair, repair of rectocele, vaginal reconstruction, or plastic surgery or repair of fistula (i.e., vesico-vaginal fistula repair, rectovaginal fistula repair, cervicouterine fistula repair, or uterine-vaginal or other fistula) *Do Not Accept* dilation and curettage (D&C), minor vaginal procedures, uterine suspension or other gynecological surgical procedures *Source:* History and physical, physician progress notes, operative notes, previous medical record if unit record
	OR			
	B. Failure to progress in labor & C-Section performed minimum 6 hours and maximum 18 hours after active phase of labor has begun	100%	None	*Failure to Progress*—in labor documented; may also be described as "failure to descend," "cervical dystocia" or "failure to dilate" *Active phase of labor*—dilation of 4 cm or greater *Source:* Labor record, operative notes, or physician progress notes
	OR			
	C. Fetal distress	100%	None	*Fetal distress*—confirmed by electronic fetal monitoring or scalp blood pH of less than 7.2 or by clinical assessment (auscultation, etc.) *Source:* Same as 1B.
	OR			
	D. Placental insufficiency	100%	None	*Placental insufficiency*—same as 1C and positive antepartum stress test (positive OCT—oxytocin challenge test) *Source:* Same as 1B and lab report
	OR			
	E. Maternal hemorrhage	100%	None	*Maternal hemorrhage*—abruptio placenta, placental abruption or placenta previa, vasco previa *Source:* Labor record, operative notes, physician progress notes or history
	OR			

	F. Malpresentations in labor or with ruptured membranes	100%	None	*Malpresentations*—breech, transverse lie, shoulder, anterior, face, brow *Source:* Labor and delivery record, history, physician progress notes, operative notes
	OR			
	G. Herpes genitalis if membranes intact or ruptured less than 4 hours	100%	None	*Herpes genitalis*—documented by culture or cytology *Source:* History, physician progress notes, laboratory and pathology reports
	OR			
	H. Feto-pelvic disproportion in labor	100%	None	*Feto-pelvic Disproportion* (FPD)—cephalo-pelvic disproportion, hydrocephaly as evidenced in X-ray, ultrasound scan or physical examination *Source:* X-ray or ultrasound reports, physician progress notes, labor and delivery record, or operative report
	OR			
	I. Maternal insulin-dependent diabetes mellitus *with* fetal maturity studies	100%	None	*Insulin Dependent*—must be taking insulin by injection or syringe *Fetal Maturity Studies*— a. Amniocentesis AND b. L/S (lecithin–sphingomyelin) ratio within 2 weeks of procedure OR c. Ultrasound anytime during pregnancy *Source:* History, physician progress notes, laboratory, X-ray, or ultrasound reports
	OR			
	J. Prolapsed cord	100%	None	*Prolapsed cord*—premature expulsion of the umbilical cord in labor before the fetus is delivered *Source:* Physician progress notes, history, Operative notes or labor record
	Patient Outcome			
2	Maternal Mortality	0%	None	
3	Mortality of infant during hospitalization	0%	None	
	Length of Stay			
4	Minimum 3 days postoperative	100%	None	
5	Maximum 7 days postoperative	100%	None	
	Other			
6	Infant discharged with mother	100%	a. Transfer to another acute care facility for neonatal intensive care or transfer to in-house intensive care unit b. Infant died c. Maternal complications	

and child that are associated with Cesarean birth. Apologists have pointed out that the rising rates have not been accompanied by an increase in overall maternal mortality and that selective use of Cesarean birth in high-risk deliveries may improve the prognosis for the child. However, consensus on indications for Cesarean birth and approaches to its optimal utilization have not been forthcoming. We report here the results of a cooperative study of 5,002 Cesarean births performed in 158 hospitals during the period from July 1, 1977 through June 30, 1978. This study was performed by physicians and staff of nine PSROs as a demonstration of the role these organizations can play in an evaluation of important issues in medical practice.

Methods

This study was conducted as a joint medical care evaluation (MCE) study by nine participating PSROs. The participating PSROs were self-selected from a group of 30 who had initially been approached by one of the authors to solicit their voluntary participation in the study as a demonstration of the potential of such cooperative studies. The criteria for the study (Table 1) were set by consensus of the physician members of the coordinating group. Draft criteria based on a brief review of the obstetric literature were initially developed by one of the authors and circulated to each of the nine PSROs.

In each area, the input of local physicians performing Cesarean deliveries (both obstetricians and family physicians) was solicited concerning the study design and criteria. A variety of mechanisms were employed by individual PSROs in obtaining such input, ranging from referral to existing PSRO and hospital medical care evaluation (audit) committees, to requesting advice from local obstetric societies, to direct mailings to hospitals and physicians. Approximately two months after the draft criteria were distributed, a physician and a staff representative from each of the nine participating PSROs met in Chicago to finalize the criteria and study design. The criteria finally agreed upon represent the consensus of physicians engaged in the practice of obstetrics in the participating PSRO areas, as reflected by their representatives on the coordinating group. Criteria were set in four categories: justification for surgery (indications), maternal outcome, length of stay and infant outcome.

The objectives of the study were to evaluate regional variations in indications, length of stay and other parameters associated with the variable incidence of Cesarean birth; to evaluate regional differences in maternal complications; and to evaluate immediate neonatal outcomes.

In each of the participating hospitals, medical records of patients undergoing primary Cesarean birth were selected chronologically, starting with patients discharged on or before June 30, 1978. All patients undergoing the procedure during the one-year period preceding that date, to a maximum of 50, were included in the study. The limitation of the sample size to a maximum of 50 records is consistent with generally accepted procedures for medical audit and was imposed to minimize the burden of data collection in any individual hospital. If more than 50 primary Cesarean births occurred, the first 50 were chosen in reverse chronologic order of discharge. More than 50 Cesarean deliveries occurred in 77 (49 per cent) of the 158 participating hospitals. Repeat Cesarean births were excluded. Additional information items were collected in each of the 158 hospitals, including the total number of deliveries, primary and repeat Cesarean births during the year; hospital size, teaching status, urban/rural location and type of ownership; use of fetal monitoring in Cesarean births; frequency of concomitant tubal ligation; and frequency of absence of prenatal care in mothers delivered by Cesarean.

Data were submitted by the hospitals to each local PSRO and were aggregated by the coordinating group. Variations from the criteria were analyzed at the hospital level, and decisions were made by peer physicians in each institution concerning the justification of those variations. Each hospital, as is normal procedure in areawide MCE studies, used either members of its own audit committee or peer physician consultants from their staff to make judgments regarding justified variations. Plans for correction of deficiencies were also developed at the hospital or PSRO level. The aggregated national data were analyzed by PSRO area.

Results

In the nine areas under study, the incidence of primary Cesarean birth during the year studied was 9.9 per cent (range 7.2 to 12.8 per cent). When repeat Cesarean births are added, the overall rate rises to 14.7 per cent (range 11.0 to 18.3 per cent). Higher total Cesarean section rates were observed in the six PSRO areas that are predominantly urban (mean 16.1 per cent, range 14.1 to 18.3 per cent) than in the three predominantly rural areas (mean 12.1 per cent, range 11.0 to 14.8 per cent). However, these differences were not statistically significant ($p > 0.05$, $df = 7$). Similarly, the rates for primary Cesarean section were also higher in urban than in rural areas, but not at a statistically significant level (urban mean 11.0 per cent, range 9.7 to 12.8 per cent, rural mean 8.1 per cent, range 7.2 to 9.9 per cent; $p > 0.10$, $df = 7$). No other patterns of geographic difference in rates were discernible. The average length of hospital stay (ALOS) ranged from 5.1 days in PSRO area F to 8.6 days in area H. The two western areas exhibited the

TABLE 2. Indications for Cesarean Section Reported in 5,002 Births in Nine PSRO Areas (% of Total Cases Meeting Indication—Totals Exceed 100% Due to Presence of Multiple Indications in Some Patients)

Indication*	PSRO Area									All Areas
	A	B	C	D	E	F	G	H	I	
1. Feto-pelvic disproportion	52.8	41.6	39.8	41.4	47.7	33.0	48.6	28.2	29.0	40.4
2. Failure to progress in labor	25.7	25.2	26.5	38.8	38.1	22.0	28.3	30.9	27.0	30.6
3. Malpresentation	22.1	28.8	21.0	30.6	32.3	19.0	27.8	21.5	30.2	26.1
4. Fetal distress	17.2	16.4	18.8	14.6	14.2	15.3	13.4	34.1	16.5	17.0
5. Maternal hemorrhage	5.4	5.6	5.7	7.4	10.2	8.0	6.7	5.6	10.2	7.2
6. Previous uterine or vaginal surgery	2.5	2.0	1.5	2.4	1.5	1.8	0.2	2.4	3.4	2.0
7. Prolapsed cord	1.3	0	1.6	2.3	1.5	2.0	1.2	2.9	2.0	1.8
8. Placental insufficiency	2.0	0.8	0.4	0.9	0.5	2.2	1.8	1.2	0.5	1.2
9. Insulin-dependent diabetes mellitus	1.1	0.4	1.0	0.7	2.0	2.0	0.8	0.9	0.5	0.9
10. Herpes genitalis	0	0	0.8	0.1	0	1.5	1.4	0	0.7	0.5
11. Other justified	4.3	6.0	7.8	4.7	3.0	18.2	6.1	0.6	NA	6.1†
12. Unjustified	0.9	0.4	0.8	1.0	3.6	1.3	1.4	0.3	NA	1.1†
Total Number of Cases	557	250	618	1537	197	600	492	340	411	5002

* Definitions for each indication are shown in Table 1.

† Does not include Area I since peer analysis of cases not meeting screening criteria was not reported in this area.

TABLE 3. Maternal and Infant Outcomes and Complications Reported in 5,002 Cesarean Births in Nine PSRO Areas

PSRO Area	Cases Studied	Maternal Mortality (per 1,000)	Infant Mortality (per 1,000)	Maternal Complications(%)*	Infant Complications(%)*
A	557	0	10.8	12.7	9.0
B	250	0	4.0	9.6	3.6
C	618	0	21.0	10.0	4.2
D	1,537	0.65†	12.4	9.3	4.1
E	197	0	20.3	1.5	3.6
F	600	1.7†	20.0	4.5	9.8
G	492	0	14.2	7.9	2.8
H	340	0	13.3	13.2	15.0
I	411	0	17.2	22.9	2.2
All areas	5,002	0.41	15.3	10.2	5.8

* Maternal length of stay greater than seven days was used as a proxy measure of maternal complications. Infant discharge later than maternal discharge was used as a proxy measure of infant complications.

† One maternal death occurred in each area.

shortest ALOS (5.1 and 5.6 days), while the longest stays were in the northeastern areas (8.6 and 7.5 days) and in one of the north central areas (8.1 days).

The indications for Cesarean birth and the frequency of each indication are shown in Table 2. Totals exceed 100 per cent because of the presence of multiple indications in some patients. Maternal and infant outcomes in each area are presented in Table 3. Postoperative length of stay of greater than seven days was used as a proxy measure for maternal complications, and infant discharge after maternal discharge served as a proxy measure of infant complications. Table 4 displays data on prenatal care, fetal monitoring and concomitant tubal ligation among the 5,002 Cesarean births studied. Area H is notable in that it exhibited an unusually low incidence of prenatal care (78.5 per cent) and a very high rate of infant complications (15.0 per cent).

Discussion

This cooperative medical care evaluation study, or medical audit, demonstrates the potential of the PSRO program for

TABLE 4. Prenatal Care, Fetal Monitoring and Concomitant Tubal Ligation Reported in 5,002 Cesarean Births in Nine PSRO Areas

PSRO Area	Cases Studied	Prenatal Care Documented(%)	Fetal Monitoring Performed(%)	Concomitant Tubal Ligation Performed(%)
A	557	97.7	56.9	7.4
B	250	100.0	59.6	10.4
C	618	91.9	62.6	8.1
D	1,537	99.7	43.5	9.8
E	197	97.5	43.7	7.1
F	600	93.5	39.7	14.2
G	492	99.6	63.8	6.1
H	340	78.5	61.8	5.9
I	411	98.3	37.0	7.8
All areas	5,002	96.1	50.4	9.0

evaluating regional differences in health care delivery. The ability to collect data from hospitals that vary in size, location, ownership and teaching status allows analysis of community health practices in a way quite different from the usual university-based health care research program. Although the cooperative study reported here is clearly quite preliminary, its success in stimulating comparison of regional practice variations has led to plans for a more rigorously designed study of technological factors affecting Cesarean birth rates, using PSROs as the data collection intermediaries.

The design of this study followed that of conventional medical audits and therefore contains several biases that necessitate cautious interpretation of the data. For instance, the self-selection of the nine participating PSROs from a larger group of PSROs, all of which we felt might be interested, raises sampling questions. Further, as is the usual practice with PSRO areawide MCE studies, hospital participation was voluntary, introducing the possibility of additional sampling bias. Nonetheless, the geographic diversity of the PSROs (Table 5) was such that we consider the data at least representative of the range of regional variation, if not representative of the total U.S. experience. Further, more than 85 per cent of the hospitals performing Cesarean deliveries in the nine PSRO areas participated in the study.

Sampling in medical audits traditionally draws consecutive records, up to a preset maximum (in our study, 50 records was the maximum in any one hospital). Of the 158 hospitals, the "sample" studied represented the entire population of Cesarean births during the one-year period in 81 hospitals. In the remaining 77, the 50 consecutive records sampled represent a potential source of nonsystematic bias, which must be considered in interpreting the results.

Despite these caveats, our data indicate that the rate of Cesarean birth continues to rise in the United States. The overall rate reported by the 158 hospitals participating in this study (14.7 per cent of all deliveries) during the period from July 1, 1977 to June 30, 1978 exceeds both the 12.8 per cent figure reported for 1977 and the 13.9 per cent projection for 1978 cited by Marieskind.[6] Approximately one third of these procedures were repeat operations, a proportion comparable to that noted in the 1974 study of PAS (Professional Activities Study) hospitals reported by Lowe.[7] So long as the incidence of primary Cesarean birth continues to increase, the pool of women who will undergo repeat operation

TABLE 5. Characteristics of the Participating PSROs

PSRO	Primarily Urban	Primarily Rural	Statewide	Geographic Area*	No. Hospitals In Study	No. Patients In Study
A	X			NE	12	557
B	X			C	5	250
C	X			SE	15	618
D		X	X	C	70	1,537
E		X		C	9	197
F		X	X	W	15	600
G	X			W	10	492
H	X			NE	7	340
I	X			C	15	411
Totals	6	3	2	-	158	5,002

* NE = Northeastern U.S.; C = Central U.S.; SE = Southeastern U.S.; W = Western U.S.

for subsequent preganancies will also continue to increase. In the 1974 PAS study,[7] only 0.9 per cent of deliveries following previous Cesarean birth were vaginal. This policy of routine repeat procedures is in contrast with the estimated 50 per cent of deliveries following previous low-segment Cesarean birth that some authors feel may be safely performed vaginally.[6]

Among the 4,591 primary Cesarean sections in our study for which peer review of cases not meeting the screening criteria was performed, only 49 (1.1 per cent) were determined by local physician peers not to be justified. Elimination of this small proportion of cases would only lower the primary Cesarean birth rate from 9.9 to 9.8 per cent, an inconsequential change. Thus, it appears that the relatively high rates of Cesarean birth being reported cannot be described as "unnecessary" surgery by contemporary local or even national standards. Rather, the rise must, in large part, be attributed to changing indications for the procedure and to changing obstetric custom.

However, the extent to which some of the criteria used in our study can be validated by scientifically sound clinical data is of concern. For instance, several of the indications included in our screening criteria are quite subjective. "Failure to progress in labor" and "fetal distress," despite our attempts to define them, remain largely matters of the judgment of the attending obstetrician. There is little question that the criteria used represent consensus of the obstetric community and are thus empirically sound descriptions of current community standards. However, one major study that attempted to determine the scientific validity of criteria developed in this manner found disappointing results.[8] This requires that our finding of only 1.1 per cent of Cesarean births not justified by peer judgment be interpreted with caution. Until such time as more scientifically valid data concerning indications for Cesarean birth are developed, consensus criteria, despite their acknowledged deficiencies, will probably continue to be used in PSRO quality assessment activities. This is a serious weakness, and careful analysis of its policy implications is required, since such criteria are used in virtually all hospital quality assurance activities.

The issues of the influence of convenience, economics and the impact of obstetric intervention itself on the rate of Cesarean birth are speculative and are likely to remain so until they are adequately investigated. In our series, "failure to progress in labor" and "feto-pelvic disproportion" were the most frequently observed indications. Some would contend, however, that many patients who undergo operations for failure to progress are in fact the products of failed induction or augmentation of labor. Whether or not the risk of delay in delivery to mother or infant in such circumstances outweighs the risk of Cesarean birth is largely an unexamined issue. Is it necessary to induce labor in all cases of prematurely ruptured membranes (PROM), or is more conservative management indicated? Kappy et al.[9] report excellent results from nonintervention in 188 women with PROM under 37 weeks of gestation. Unfortunately, our study did not examine the relationship between PROM and Cesarean delivery. This issue, however, will be the subject of further investigation, as will the timing of the choice between Cesarean and vaginal delivery in mothers with PROM.

Oxytocin administration for induction or augmentation of labor may be a major factor in the increased rate of Cesarean births.[10] Difficulty in maintaining adequate control of contractions can produce uterine tetany, induce fetal distress, or lead to early fatigue and arrested labor; conversely, inadequate trials of oxytocin may lead to a premature decision for Cesarean birth. Oxytocin administration is certainly a valuable therapy when medical indications for induction or augmentation of labor exist. However, the dictum of *primum non nocere* must be a prime consideration in making decisions regarding its use. A portion of the increase in incidence of Cesarean birth may be related to failures of elective induction of labor. Further research is needed to evaluate this hypothesis. We believe, however, that elective induction of labor for physician convenience, or solely for presumed fetal postmaturity, has no place in contemporary obstetrics.

Dystocia (including both feto-pelvic disproportion and failure to progress in labor) has accounted for more than one third of the rise in Cesarean birth rates.[3] The goals of "undamaged infants and reasonable duration of labor" are cited by Bottoms et al.[3] as major reasons for this change. A substantial proportion of the obstetric literature[11–14] supports the contention that more liberal use of Cesarean delivery is associated with improved infant outcome. Clinical data to support this conclusion, however, are almost nonexistent. Until adequate clinical trials can show causal relationship between neonatal mortality and morbidity and route of delivery, Cesarean birth is certain to continue to come under attack. We do not intend by this discussion to underestimate or diminish the value of Cesarean birth in contemporary obstetric practice. Rather, we urge that adequately controlled trials be performed to assess and document the benefits and costs associated with it in management of specific complications of pregnancy and birth.

Electronic fetal monitoring has also been implicated as a factor in the rising incidence of Cesarean birth. Haverkamp reported a doubling of the Cesarean birth rate in monitored patients with no improvement in perinatal outcomes.[15] Banta and Thacker concluded from a review of the literature that electronic fetal monitoring (EFM) "is no better than auscultation" and that it "adds little except to increase the Cesarean section (CS) rate."[16]

Conversely, Evrard and Gold felt that "the preponderance of evidence shows that fetal monitoring is predictive of fetal hypoxia, will slightly increase the CS rate, and will decrease the perinatal mortality."[17] Until such time as adequately controlled, randomized trials have been performed to assess the value of fetal monitoring in predicting perinatal results and to compare its sensitivity and selectivity with that of less invasive techniques, the debate will continue. In our study, we found no association between the Cesarean birth rate in the PSRO area and the incidence of patients who underwent Cesarean birth previously being electronically monitored. Unfortunately, however, we did not collect data to indicate the total proportion of pregnancies in which fetal monitoring was used. This issue will also be a major focus of subsequent study.

Maternal outcomes in our study were good. Two deaths occurred in the 5,002 cases studied, a mortality rate of 40/100,000. Although this is certainly a reflection of the relative safety of the operation, it is more than three times higher than the 1976 U. S. maternal mortality statistic of 12.3/100,000. If one assumes the 40/100,000 statistic we found to be a component of the 12.3/100,000 figure for both routes of delivery, the relative mortality risk of Cesarean birth is more than five times that of vaginal delivery. This speaks to the necessity of assuring that Cesarean delivery is performed only where clearly indicated.

The other measure of maternal outcome we studied was complication rate. Using a hospital stay of greater than seven days as a proxy for complications, we found that an average of 10.2 per cent of the patients had "complications" that extended their hospitalization. However, much of the local variation in this statistic is an artifact of the well known, and unexplained, difference in average length of stay between the eastern and western parts of the United States. Thus, this finding too must be interpreted cautiously.

Infant outcomes were assessed in our study by the proxy measure of discharge with or before the mother. Although the range was large (2.2 to 15.0 per cent), the average infant complication rate (5.8 per cent) was reasonably low, considering that Cesarean birth is generally reserved for high-risk infants. The overall infant mortality in our series (15.3/1,000) is not appreciably different from the 1976 U.S. perinatal mortality ratio (17.3/1,000).

We were also pleased to note that the vast majority of women we studied (96.1 per cent) had received at least some prenatal care. In only one area (H) did as many as 20 per cent of the women who underwent Cesarean birth have no documentation of prenatal care. However, the rate of infant complications in area H (15.1 per cent) was significantly higher ($p < 0.05$, $df = 7$) than in the other eight areas. This reinforces the importance of adequate prenatal care, especially in high-risk pregnancies.

Conclusions

This study demonstrates the viability of using local PSROs as a means of conducting national studies on important issues in medical practice. Despite a number of design and sampling weaknesses that limit the strength of our conclusions, our data demonstrate that, although the rate of Cesarean births continues to rise, only an inconsequential number of these operations are inappropriate in the judgment of peer physicians in the participating hospitals. However, the relative risks and benefits of the higher rates have not been adequately examined. Further study of such concerns as the influence of oxytocin induction, economics and physician convenience, and electronic fetal monitoring on the rising rate of Cesarean birth is essential to define the proper role of this form of delivery in contemporary obstetrics. The PSRO may be a valuable vehicle for conducting such studies on a communitywide basis.

References

1. Goran MJ, Roberts JS, et al. The PSRO hospital review system. Med Care 1975;13(Supplement):1-33.

2. Jessee WF, Munier WB, et al. PSRO: an educational force for improving quality of care. N Engl J Med 1975;292:668.

3. Bottoms SF, Rosen MG, Sokol RJ. The increase in the Cesarean birth rate. N Engl J Med 1980;302:559.

4. Obstetrical practices in the United States, 1978: Hearing before the Subcommittee on Health and Scientific Research of the Committee on Human Resources, United States Senate. Washington, D.C.: Government Printing Office, 1978.

5. Pettiti D, Olson RO, William RL. Cesarean section in California—1960 through 1975. Am J Obstet Gynecol 1979;133:391.

6. Marieskind HE. An evaluation of Cesarean section in the United States. Washington, D.C.: Office of the Assistant Secretary for Planning and Evaluation/Health, Department of Health, Education and Welfare, 1979.

7. Lowe JA, et al. Cesarean sections in U.S. PAS hospitals. PAS Reporter 1976; Vol. 14, December.

8. Respess J. Application of process and outcome measurements in an EMCRO. In: Evaluation in health services delivery. New York: Engineering Foundation Conferences, 1973.

9. Kappy KA, Cetrulo CL, et al. Premature rupture of the membrances: a conservative approach. Am J Obstet Gynecol 1975;134:655.

10. Dunn PM. Obstetric delivery today: for better or for worse? Lancet 1976;i:790.

11. Pakter J, Nelson F. Factors in the unprecedented decline in infant mortality in New York City. Bull NY Acad Med 1974;50:839.

12. Diddle AW, Senamer JR, Slowey JF. Cesarean section: changing philosophy. Postgrad Med 1973;53(3):160.

13. Sutherst JR, Case BD. Cesarean section and its place in the active approach to delivery. Clin Obstet Gynecol 1975;21:241.

14. Paul RH, Huey JR Jr, Yaeger CF. Clinical fetal monitoring: its effect on Cesarean section rate and prenatal mortality: five year trends. Postgrad Med 1977;61(4):160.

15. Haverkamp A, Thompson H, et al. The evaluation of continuous fetal heart rate monitoring in high-risk pregnancy. Am J Obstet Gynecol 1976;125:310.

16. Banta HD, Thacker SB. Costs and benefits of electronic fetal monitoring: a review of the literature. Hyattsville, MD: National Center for Health Services Research, 1979. (DHEW Publication No. (PHS)79-3245.)

17. Evrard JR, Gold EM. Cesarean sections: risk/benefit. Perinatal Care 1978;2(8):4.

MEDICAL CARE
January 1982, Vol. XX, No. 1

Periodic Medical Review: The Impact in Pennsylvania

MICHAEL A. CALOGERO, M.S., NHA,* AND DANIEL B. HILL, PH.D.†

Periodic Medical Review (PMR) of all Medicaid beneficiaries in skilled nursing facilities (SNFs) in Pennsylvania was implemented in fiscal year 1973-74. In this study, the authors explored the impact of PMR through its first four annual review periods. Because over two thirds of all SNF patients who were determined (over the four-year period) to require lower levels of care were identified during the first year of the program, the study was focused on the impact of first-year reviews on patient placements and nursing home industry characteristics in the subsequent years. Although the results did not strongly show that PMR had significant long-term effects, the general conclusions reached were that: 1) first year PMR decisions altered (in the intended direction) SNF patient placements and structural characteristics of the nursing home industry in Pennsylvania; 2) the initial impact of PMR was still apparent by the fourth year, but much weaker; and 3) the PMR review team may have been more lenient in their decisions concerning appropriate levels of care for reviewed SNF patients after the first year.

DURING THE PAST DECADE, expenditures for nursing home care have been the fastest growing component of national health expenditures. From a 1965 level of about $2 billion, nursing home expenditures increased to almost $16 billion in 1978.[1] Expenditures by federal, state and local governments represented 53 per cent of nursing home expenditures in 1978[1]; federal and state expenditures under the Medicaid program were 87 per cent of all government expenditures for nursing home care.[1] Medicaid spending for nursing home care constituted almost 40 per cent of all Medicaid expenditures during 1978.[1] These data illustrate the importance of nursing home costs as a public policy issue and, in particular, the financial implications for the Medicaid program.

A number of researchers have concluded that significant portions of nursing home patients could receive appropriate care at lower levels of institutional care or in community settings if such care were available.[2-7] These findings raise concern regarding the use of Medicaid funds, but they also have implications for the inappropriately placed patient. The provision of unnecessary services may lead to physical, motivational and emotional problems for the elderly patient.[8-13]

Periodic Medical Review

In order to monitor the levels of care at which nursing home patients are placed

---

* Administrator, Olivette Skilled Nursing Care Center, Elgin, Illinois.

† Associate Professor, Department of Health Services Administration, School of Community and Allied Health, University of Alabama in Birmingham.

Address for reprints: Michael A. Calogero, M.S., NHA Administrator, Olivette Skilled Nursing Care Center, Elgin, IL 60120.

0025-7079/82/0100/0085/$01.10 © J. B. Lippincott Co.

and to determine the appropriateness of care given to Medicaid-funded patients, Congress mandated a program of Periodic Medical Review (PMR) in 1969. PMR is a state-run program involving external peer review of the care rendered Medicaid patients in skilled nursing facilities and mental hospitals.

> The primary goal of Periodic Medical Review is to ensure the provision of a range and quality of medical and nursing management, and social work support for Title XIX skilled nursing home and mental hospital patients that is both necessary and commensurate with their clinical and physical needs and optimum social functioning.[14]

Under PMR, each skilled nursing facility (SNF) that serves Medicaid patients is to have an on-site inspection by a Medical Review (MR) team at least annually. Federal regulations stipulated that the Medical Review team be composed of one or more physicians and other appropriate health and social service personnel.‡ In Pennsylvania, during the time period studied, each team consisted of one physician, one or more registered nurses and one social worker.

An inspection involves a separate review of each patient's medical records by a representative of each professional discipline on the team. Direct observation and interview (if possible) of each patient also contribute to the determination of appropriateness of placement. Group discussion of all cases results in determination of the number of and reasons for inappropriate placements. The findings of the team are presented to the administration of the facility in an exit interview, along with a description of the necessary action for correction of any deficiencies and suggestions for improvements in care and/or services to patients. Once the team decisions are presented, the patient and/or the physician have the option to appeal. If neither appeal or if the appeal is not successful, then the patient must be moved to the level of care stipulated by the MR team. This may involve a simple reclassification or a physical relocation of the patient.

Issues Concerning PMR

The potential effectiveness of Periodic Medical Review of SNF patients is affected by two major sets of issues. The first relates to the nature of the long-term care patient and his/her needs. The second set of issues involves the more general concerns raised when governmental agencies attempt to regulate the behavior of private firms.

Appropriate services for an elderly long-term care patient are much different from those for an acute-care hospital patient. An acute-care hospital is a disease-oriented facility with its resources focused on the disease rather than on the patient.[15] In contrast, the nursing home patient is usually institutionalized for an extended period of time and is dependent upon the institution for a wide range of medical, social and psychological needs.[16] Because the needs of the long-term care patient cannot be stated in strictly medical terms, it is difficult to specify standards of appropriate care that are applicable to the individual patient. There exists a substantial lack of knowledge as to what constitutes quality and/or effective care in the long-term care setting.[17,18] While a number of rating scales and patient classification schemes have been developed, definitive standards for use by external reviewers are not available.[3,10,19–21] Based on the desirability of considering the multiplicity of patient needs, the interdisciplinary team approach to medical review has received considerable support.[22,23]

As a regulatory mechanism charged with making judgments in the absence of well-defined standards, PMR has the potential to be afflicted with the problems that have been associated with other regulatory programs. A number of authors have argued that a regulatory program will tend to have a declining level of effectiveness over the life cycle of the program and may eventually benefit the regulated parties.

---

‡ A physician is no longer required as part of the on-site team, but was during the time period addressed in this study.

As Jaffee stated, "The American people have no enthusiasm for regulation. In the presence of general breakdown or glaring evils they will respond with regulation, but their concern will not survive the resolution of the crisis."[24] While the "people" may lose interest in a regulatory program, those parties most directly affected (the regulated industry) will focus their energies on trying to influence the nature and scope of regulation.[18,24,25] Faced with declining public support, sometimes evidenced by inadequate funding,[25] and cognizant of the fact that they might be blamed for any failures that occur in the regulated industry,[24,26] the regulators can become captive to the interests of the regulated industry.[24-27] This outcome, sometimes attributed to the self-preservation instincts of the regulatory agency,[24-27] is most likely to occur when the public interest that the agency is intended to serve is difficult to assess.[26]

It is not obvious whether a pessimistic prognosis for the effectiveness of Periodic Medical Review is warranted. A scenario depicting the capture of this regulatory program by the regulated industry is plausible. The program was initiated during a time of great concern over nursing home abuses and the care rendered to the elderly.[28-30] However, in contrast with many regulatory programs, PMR is charged with more than the protection of the general public interest. The program is administered by agencies (the state Medicaid agencies) that have a direct financial stake in the effectiveness of the program.

Because the study reported herein did not involve observations of the Medical Review decision-making process, the capture hypothesis was not directly tested. The authors examined the impact of PMR on patient placement patterns aggregated at the county and state levels in Pennsylvania. While the results are *consistent* with the existence of some degree of nursing home influence on the regulatory process, to interpret the findings as supportive of the capture hypothesis would be highly speculative.

Other Studies

Two prior studies have examined the recommendations of Medical Review teams under the PMR program. A Rhode Island study reported the results of Medical Review of SNF patients during the program's first year, 1973. Of the Medicaid SNF patients reviewed, 30.3 per cent were determined to be inappropriately placed. Of the inappropriate placements, 97 per cent were recommended for transfer to intermediate care facilities (ICFs), and 3 per cent were recommended for community care.[6]

During 1975, the PMR program in Massachusetts identified 18.2 per cent of all Medicaid SNF patients in the state as being inappropriately placed. The number and per cent of patients so identified varied considerably between facilities.[7]

Each of these two studies examined one dimension, review team decisions, of the PMR program during one early year of the program. While the magnitude of the inappropriate placements reported is of interest, the studies provide only limited insight concerning the effectiveness of PMR in stimulating a more efficient use of Medicaid funds. In the study reported herein, the authors attempted to capture a more complete picture of the impact of PMR in one state, Pennsylvania.

Data

The time period covered by this study corresponds to the first four annual PMR periods in Pennsylvania: fiscal years (July through June) 1973-74, 1974-75 and 1975-76, and calendar year 1977. Data on the number of SNF patients reviewed and review team recommendations were derived from summary work sheets provided by the Division of Medical Review, Pennsylvania Department of Public Welfare. Due to the Department's concern for the confidentiality of patient and facility data, totals were compiled on a county-by-county basis. No individual facility data were used. Pennsylvania has 67 counties, although only 65 had nursing homes during the time period studied. Thus, there were 65 separate totals used in this study.

Other data collected included elderly population (65 years or over) by county,[31] percentage distribution of urban and rural elderly residents by county,[32] and SNF beds and ICF beds by county.[33,34]

Hypotheses

The direction pursued in this study was based on initial observations of aggregated data for the state of Pennsylvania. As shown in Table 1, the per cent of Medicaid SNF patients who were recommended for alternative levels of care varied drastically between the initial year of PMR (1973-74) and subsequent years. The per cent of patients recommended for change in care level in the first year (34.05 per cent) exceeded the corresponding figures for the Rhode Island and Massachusetts studies.[6,7] In contrast, recommended changes accounted for no more than 6.1 per cent of patients reviewed in each of the second through the fourth years.

Based on these observations, three hypotheses were developed. Each of the hypotheses, which appear below, is consistent with the significant reduction in recommended changes in level of care following the first of PMR.

> H1. The Medical Reviews of SNF patients in 1973-74 resulted in a major relocation of SNF patients to lower levels of care and had a lasting impact on the SNFs Medicaid patient placement and retention practices in subsequent years.
>
> H2. While the Medical Reviews of SNF patients during the first year resulted in a significant relocation of SNF patients to lower levels of care, SNFs returned to their pre-PMR patient placement and retention practices after the second-year reviews.
>
> H3. PMR had no significant impact on SNF patient placement in either the short run or the long run (null hypothesis).

If H1 were valid, it could be implied that, as a result of the first year of experience with PMR, the SNFs improved their in-house review mechanisms for monitoring patient placements. If either H2 or H3 were valid, the implication is that the reduced number of recommended changes after the first year reflected modifications in the PMR criteria. Several reasons for review criteria modifications can be postulated, although it was not possible to test the validity of these explanations in this study. First, the PMR teams may have determined that first-year reviews placed too much weight on medical criteria with in-

TABLE 1. Comparative Results of Four Medical Reviews in Pennsylvania

	1973–74	1974–75	1975–76	1977
Patients reviewed (total)	18,046	16,402	19,797	23,140
Patients recommended for change in level of care (total)	6,145	869	1,225	758
% of total recommended for change in level of care	34.05	5.30	6.19	3.28

Source: Pennsylvania Department of Public Welfare, Division of Medical Review, unpublished data obtained through personal contact, 1978–79.

sufficient consideration of psychological and social factors. Second, PMR teams may have felt that the first-year reviews were not sensitive to the availability of alternative levels of care. Third, the PMR team members may have become more sensitive to the viewpoints of SNF personnel regarding patient placement. H3 carries the additional inference that the PMR program had little success in having recommended changes actually occur, either because of weak enforcement and/or as a result of successful appeals by SNFs. Data concerning compliance with PMR recommendations were not available for this study.

Another potentially important variable would be the change in review team personnel that might have occurred over the four years. Although precise information was not available, officials of the Department of Medical Review stated that there was not a significant turnover of review team personnel during this time period.

Methodology

Because 68.3 per cent of all recommended changes for Medicaid SNF patients during the first four annual reviews occurred during the first year, the per cent of Medicaid SNF patients recommended for change in level of care during the first year (CHANGE), by county, was the variable representing the focal point of this study. This variable was correlated with supply measures for the first year, and with patient population and supply measures for the second year (to determine initial impact) and the fourth year (to determine lasting impact).

The impact of PMR on nursing homes in Pennsylvania was examined by determining the rank correlations between CHANGE and two types of variables: Medicaid SNF patients and structural variables. The tau b statistic was used to measure the direction and strength of correlations.

Medicaid SNF Patients

Because all Medicaid SNF patients are reviewed annually under PMR, the number of reviews (PTSREV) during a review year provides a measure of the average Medicaid SNF patient population for the year. Use of this measure carries the implicit assumption that the reviews are evenly spaced over the year. While a number of factors might affect the number of Medicaid beneficiaries in SNFs across the state, it was felt that if PMR had a significant impact on patient placements, the *relative changes* in Medicaid SNF patients across counties during the study period should be correlated with PMR recommendations. $PTSREV_1$, $PTSREV_2$, and $PTSREV_4$ connote Medicaid SNF patients reviewed in the programs during the first, second and fourth years, respectively; $PTSREV_2/PTSREV_1$ is the ratio of second-year reviews to first-year reviews, etc.

Structural Variables

Medicaid patients occupied 40 per cent of all SNF beds in Pennsylvania during PMR's first year; this figure was 48 per cent during the program's fourth year (1977). Medicaid SNF patients recommended for lower levels of care in 1973-74 occupied 13.6 per cent of all SNF beds in the State. These figures indicate that if PMR were successful in altering patient placements, some correlation between CHANGE and subsequent growth rates for SNF beds might be anticipated. An even more pronounced (relative) impact on ICF beds could be anticipated, due to the fact that there were almost four SNF beds per ICF bed during the first year. The number of SNF patients recommended for lower levels of care in the first year, over 99 per cent of whom were recommended for ICF placement, equaled 52.7 per cent of the number of existing ICF beds.

TABLE 2. Correlations Anticipated Under Three Hypotheses; Dependent Variable: Change

Variable	Hypotheses		
	H1	H2	H3
$\text{PTSREV}_2/\text{PTSREV}_1$	–	–	N
$\text{PTSREV}_4/\text{PTSREV}_2$	N	+	N
$\text{PTSREV}_4/\text{PTSREV}_1$	–	N	N
$\text{SNFBDS}_2/\text{SNFBDS}_1$	–	–	N
$\text{SNFBDS}_4/\text{SNFBDS}_2$	–^{a}	+	N
$\text{SNFBDS}_4/\text{SNFBDS}_1$	–	N	N
$\text{ICFBDS}_2/\text{ICFBDS}_1$	+	+	N
$\text{ICFBDS}_4/\text{ICFBDS}_2$	+^{a}	–	N
$\text{ICFBDS}_4/\text{ICFBDS}_1$	+	N	N

– means negative correlation anticipated.
+ means positive correlation anticipated.
N means no significant correlation anticipated.
^{a} means correlation anticipated if lag time was necessary for full impact on bed supplies.

The supplies of SNF beds and ICF beds in each county, measured on a per-100-elderly (65+) population basis, are represented by SNFBDS and ICFBDS, respectively. (In 1975, less than 3 per cent of Pennsylvania nursing home patients were under the age of 65.[35])

TABLE 3. Tau b Values Measuring Correlations of Selected Variables with Change

Variable	Tau b Value
SNF Beds/100 Elderly Population	0.16477*
ICF Beds/100 Elderly Population	−0.05682
Hospital Beds/1,000 Total Population	−0.10440
$\text{PTSREV}_2/\text{PTSREV}_1$	−0.32955†
$\text{PTSREV}_4/\text{PTSREV}_2$	0.08062
$\text{PTSREV}_4/\text{PTSREV}_1$	−0.12074
$\text{ICFBDS}_2/\text{ICFBDS}_1$	0.04372
$\text{ICFBDS}_4/\text{ICFBDS}_2$	0.08668
$\text{ICFBDS}_4/\text{ICFBDS}_1$	0.14988*
$\text{SNFBDS}_2/\text{SNFBDS}_1$	−0.04121
$\text{SNFBDS}_4/\text{SNFBDS}_2$	0.02137
$\text{SNFBDS}_4/\text{SNFBDS}_1$	−0.02060

* Statistically significant at the 0.1 level.
† Statistically significant at the 0.001 level.

Two considerations should be raised regarding the relationships between CHANGE and structural variables. First, the relationships might not be expected to be as strong as that between CHANGE and PTSREV. Some modifications in Medicaid patient placement might be accommodated through altered levels of excess capacity and changes in non-Medicaid patient census. Second, one might question the impact that PMR could have on nursing home construction during the time framework covered by this study (roughly 3½ years after the first review). However, the possibility exists that some nursing homes could alter the number of beds in existing facilities, while (more importantly) others might reclassify beds from SNF to ICF.

The results to be anticipated under each of the three hypotheses are presented in Table 2. In general, the correlation of CHANGE with the ratio of the values of a variable in the first two years is used to represent the initial impact of PMR; an opposite direction of correlation with the fourth year to second year ratio represents a reversal of the initial impact; and a fourth year to first year ratio sharing a similar correlation, as does the ratio for the first two years, represents a retention of the initial impact. The ratio of SNF Medicaid patients reviewed between two years is used as an indication of the relative impact of PMR across counties. This implies that factors other than PMR that might have affected SNF patient population trends had uniform or, at least, random effects across counties.

The tests represented in Table 3 reflect the major thrust of this study. Additional tests, conducted to gain further insight from the available data, are reported under in the "Additional Analyses" section of this article.

Results

The results derived from testing the three hypotheses are summarized in Table

3. Table 4, which will be used to complement the interpretation of Table 3, provides relevant data for the state.

The most direct impact of first-year PMR on patient placement would be reflected in SNF Medicaid censuses in the second year. In both the state data and the test using county data, the findings support the conclusion that PMR affected the placement of SNF patients. For the state, the SNF Medicaid population decreased by 9.11 per cent between 1973-74 and 1974-75 (Table 4). The fact that the SNF Medicaid population was 25.72 per cent *higher* in 1977 than during 1973-74 indicates that the second-year reduction was counter to the trend, lending support to the conclusion that the second-year figure demonstrates a response to first-year PMR decisions. The correlation of CHANGE with PTSREV$_2$/PTSREV$_1$, using county data, provides further support for this conclusion (Table 3). The relationship is in the anticipated direction (negative) and significant at the 0.001 level.

The findings concerning structural variables in the first two years of PMR are consistent with H1 and H2. As shown in Table 4, there was a sizable increase in ICF beds (38.74 per cent) in the second year, while SNF beds increased by a modest amount (2.88 per cent) in the second year. The tested correlations of CHANGE with the second year to first year bed ratios produced relationships in the anticipated direction, but these relationships were very weak and not statistically significant.

Comparative analyses of data for year 4 (calendar year 1977) with data for years 1 and 2 provided tests of H1 versus H2. H1 predicts a retention in year 4 of the impact of first-year reviews as witnessed in year 2; H2 predicts a dissolution of that impact. The findings do not provide a firm basis for

TABLE 4. Change in Selected Variables—Years 1, 2, and 4
Total Group—All Counties

Selected Variables	Year 1 (1973–74)	Year 2 (1974–75)	Year 4 (1977)
Patients reviewed* (total)	18,046	16,402	23,140
% change from year 1	—	− 9.11	25.72
% change from year 2	—	—	41.08
SNF Beds† (total)	45,084	46,381	48,160
% change from year 1	—	2.88	6.82
% change from year 2	—	—	3.84
ICF Beds† (total)	11,665	16,180	15,694
% change from year 1	—	38.70	34.54
% change from year 2	—	—	− 3.03

---

* Pennsylvania Department of Public Welfare, Division of Medical Review. Unpublished data obtained through personal contact, 1978–79.

† Pennsylvania Department of Public Welfare. Unpublished data obtained through personal correspondence, Februrary 1979.

accepting one hypothesis over the other. The correlations between CHANGE and $PTSREV_4/PTSREV_1$, $SNFBDS_4/SNFBDS_1$ and $ICFBDS_4/ICFBDS_1$ are all in the direction anticipated under H1, but only the ICF variable provides a relationship that is statistically significant at the 0.1 level. For the other two variables, the correlations between CHANGE and the fourth year to second year measures are in the opposite direction but weaker than the second year to first year correlations. From Table 4, it can be seen that, as in year 2, the relative growth in the number of ICF beds between years 1 and 4 was much greater than the growth in SNF beds for the state. All of the growth in the number of ICF beds occurred between years 1 and 2.

Additional Analyses

Three additional types of analyses were conducted in an attempt to shed additional light on the findings. First, tests were conducted using subsamples of counties, urban and rural. Second, the correlation between CHANGE and $CHANGE_2$, the corresponding measure of PMR decisions in the second year, was determined. Finally, correlations between CHANGE and nursing home bed supplies during the first year were examined.

Urban/Rural

The counties in Pennsylvania were classified into two groups, urban and rural. Those counties having 55 per cent or more of their elderly population defined as rural[36] were classified as rural counties. The dividing line was chosen to provide equivalent numbers of counties in each subsample.

The tests conducted for the original sample were again performed for each subsample. The purpose of this analysis was to determine whether the subsamples could reinforce the findings derived from the larger sample. A second rationale for examining the two subsamples was provided by the findings of Davis and Gibbin.[37] In a study of nursing home residents in the Buffalo, New York area, Davis and Gibbin determined that the percentage of urban patients judged to be inappropriately placed was twice that of rural patients. They attributed the more common misuse of urban facilities to the less extensive social support network (family and friends) characteristic of the urban elderly population. The subsample findings are summarized below.

1. The most direct test of the initial impact of PMR is the correlation between CHANGE and $PTSREV_2/PTSREV_1$. The negative correlation observed for the total sample was substantiated for the urban subsample (−0.258 tau b, 0.05 significance level) and for the rural subsample (−0.414 tau b, 0.01 significance level).
2. The test of the rank correlation between CHANGE and $PTSREV_4/PTSREV_1$ indicated some retention of the initial PMR impact in rural counties (−0.251 tau b, 0.1 significance level), but no retention in urban counties (0.0 tau b).
3. The finding of a positive correlation between CHANGE and $ICFBDS_4/ICFBDS_1$ in the total sample was attributable only to the rural counties (0.387 tau b, 0.01 significance level).
4. Both subsamples were consistent with the total sample in a lack of significant correlations between CHANGE and the SNF beds variables.

In general, tests applied to the total population and to each of the subsamples support the rejection of H3, the null hypothesis. The findings for urban counties were consistent with H2. The findings for rural counties supplied some support for H1.

The lack of a retained impact on patient placements in urban SNFs might have resulted from two factors. First, social support considerations, as noted by Davis and Gibbin, may have assumed increased importance in the placement of urban patients. Second, the finding that urban ICF bed supplies were not responsive to PMR decisions indicates a lack of institutional alternatives for the urban SNF patients.

Second-Year Recommendations

The rank correlations between recommended changes in patient placements during the first and second year of PMR were examined for two reasons. First, it was felt that the consistency of review decisions might be reflected in this relationship. Second, it was recognized that the weaker associations between CHANGE and SNF Medicaid patients in the fourth year might be attributed to a reversal of the comparative levels of recommended changes across counties after the first year.

To address directly the second point above and to provide an initial test of the first point, the relationship between CHANGE and CHANGE_2 was tested. Because of the considerable difference between the per cent of SNF patients recommended for relocation in the first year (34.05 per cent) and the reduction in SNF patients between the first and second year (9.11 per cent), it was felt that consistency in the review criteria would be reflected in a positive relationship between CHANGE and CHANGE_2. On the other hand, a negative relationship between CHANGE and CHANGE_2 would provide an explanation for the weaker relationship between CHANGE and the patient placement variable ($\text{PTSREV}_4/\text{PTSREV}_1$) for the fourth year. The observed correlation between CHANGE and CHANGE_2 (0.06300 tau b) was insignificant, providing support for neither consistency in review criteria nor a reversal of the relative impact of first-year reviews. No significant correlation was observed in either the urban county subsample or the rural county subsample.

A second test of the correlation between first-year and second-year reviews was conducted to provide further analysis of the apparent consistency of review criteria. For this test, a variable was constructed to represent the *expected* number of Medicaid SNF patients reviewed in year 2 (EPTSREV_2). For each county, the expected reduction in patients reviewed was computed by multiplying the number of patients recommended for change in the *county's* SNFs during the first year (CHANGEREC) by the *state* ratio of second-year reduction in SNF patients reviewed (1,644) to first-year recommended changes (6,145), i.e., $\text{EPTSREV}_2 = \text{PTSREV}_1 - (\text{CHANGEREC}) \times (1{,}644/6{,}145)$. A value of actual second-year patients reviewed in excess of EPTSREV_2 for a county would indicate that the SNFs in that county had altered patient placements, relative to the level of recommended changes, to a lesser extent than had SNFs across the state. Thus, if the PMR review criteria were consistent in the first two years, one would expect a positive correlation between CHANGE_2 and $\text{PTSREV}_2/\text{EPTSREV}_2$. On the other hand, a negative correlation would indicate a change in review criteria during the second year and would suggest that the relative impact across counties of PMR decisions in the first year might have been reversed after the second year. In testing this correlation, no relationship was found between $\text{PTSREV}_2/\text{EPTSREV}_2$ and CHANGE_2 (−0.00073 tau b), a finding also observed for the urban and rural subsamples.

In summary, the tests linking PMR reviews in the first two years provide no evidence that second-year reviews *directly* modified the impact of first-year reviews on SNF patient placements. However, it appears that second-year review decisions were not affected by the responsiveness of SNFs to recommendations made during the first review period. This suggests that more lenient review criteria were applied in the second year. If this were the case, it might indicate that a) PMR criteria in the first year were unrealistically narrow and applied too rigidly and/or b) the PMR program was less effective in identifying inappropriately placed SNF patients during the second year.

Change and Bed Supplies

The alternative explanations presented above (a and b) bring to light a critical consideration in judging the PMR program after its initial year. If one is to judge the effectiveness of the program in its first year by its impact on patient placements, then one must assume that the program effectively identified *inappropriately* placed patients. This assumption cannot be tested directly with available data. An indirect assessment can be made by correlating first-year decisions with variables that are considered to be determinants of inappropriate utilization. Several authors have concluded that "Reomer's law," an availability effect, exists with respect to the utilization of inpatient hospital care.[38-40] According to these researchers, the demand for inpatient care is positively correlated with the available supply of beds, and patients for whom the need for treatment in an acute care facility is questionable will represent a greater portion of patient populations in areas with a relative abundance of beds. The logic behind the availability effect can be extended to the nursing home bed supplies. However, a study by Willemain and Farber[41] provided somewhat tentative findings that contradicted the existence of an availability effect with respect to inappropriate placement of patients in SNFs. Their study used PMR decisions concerning the appropriateness of SNF placements in eight health planning regions in Massachusetts. Controlling for ICF beds and hospital beds, they observed a negative correlation between SNF beds (per 1,000 elderly) and inappropriate placements in SNFs. Second, they found an inverse correlation between utilization of SNFs and the supply of ICF beds.

In this study, the relationships between CHANGE and nursing home bed supplies in 1973-74 in Pennsylvania were tested. The bed supply variables were specified as beds per 100 elderly (65+) residents. The results did not indicate strong correlations, but the observed correlations were consistent with the existence of availability effects and contrary to the findings of Willemain and Farber. A positive correlation was observed between CHANGE and SNF beds (0.165 tau b, 0.1 significance level); a negative correlation was observed between CHANGE and ICF bed supplies (−0.057 tau b, not significant at the 0.1 level). The direction of these relationships held for the urban and rural subsamples, with the rural subsample having a stronger CHANGE–SNF relationship and the urban subsample having a stronger CHANGE–ICF relationship.

These results provide support, albeit indirect and somewhat weak, for the conclusion that SNF patients identified as being inappropriately placed by first year PMR reviews did include a more than proportionate share of all Medicaid SNF patients who were "inappropriately" placed in SNFs. It should be noted that this conclusion does not negate the possibility that first year review criteria were too rigid or applied too aggressively, e.g., that the reviewers were not sensitive to social or emotional variables.

Conclusions

Periodic Medical Review of all Medicaid beneficiaries in skilled nursing facilities in Pennsylvania was implemented in fiscal 1973-74. In this study, the authors explored the impact of PMR through its first four annual review periods. Because over two thirds of all SNF patients who were determined (over the four-year period) to require only lower levels of care were identified during the first year of the program, the study was focused on the impact of first-year reviews on patient placements and nursing home industry characteristics in the subsequent years. In assessing the conclusions listed below, the reader should bear in mind the qualifications raised previously concerning the lack of statistical significance for many of the tests. Three major conclusions are offered in descending order of the strength of the tests upon which they are based.

1. First year PMR decisions altered (in the intended directions) SNF patient placements and structural characteristics of the nursing home industry.

This conclusion is supported by aggregate data for the state and cross sectional data using county units. The first year reviews stimulated the relocation of SNF patients (and/or modifications in admissions practices) and the growth in ICF bed supplies relative to SNF bed supplies.

2. The initial impact of PMR decisions was still apparent, but weaker, during the fourth annual review period.

The declining impact of PMR after its second year raises questions concerning the long run cost effectiveness of the program.

3. PMR teams were more lenient in their appropriateness decisions concerning SNF patients after the first year of the program.

The evidence does not support the hypothesis that the drastic decline in relocation recommendations after the first year can be totally explained by reformed behavior by SNFs regarding the placement of Medicaid patients. It cannot be determined whether the greater leniency reflected the adoption of more appropriate review criteria. However, the apparent lack of association between second-year review decisions and the responsiveness of SNFs to first-year decisions lends support to the hypothesis that the effectiveness of the program was reduced after the first year.

Ultimately, the effectiveness of Periodic Medical Review rests on its success in accomplishing (or stimulating) three results:

1. The identification of inappropriately placed patients;

2. The relocation of these patients to appropriate settings; and

3. The adoption by SNFs of appropriate admission and discharge policies.

While evaluation of the PMR program by the first criterion represents an imposing research challenge, the second criterion is more amenable to empiric inquiry. The concerns raised by this study regarding the long-term effectiveness of PMR provide strong support for the advisability of tracking SNF patients who have been recommended for relocation to determine their subsequent distribution by loci of care. Knowledge concerning the placement of patients recommended for relocation during PMR's first year, the reasons behind cases where relocation did not occur and the nature of PMR decisions regarding these patients in subsequent years could greatly enhance the initial insights produced by the study reported herein.

References

1. Gibson RM. National health expenditures. Health Care Finance Review 1979;4:1.

2. Davis JJ, Gibbin MJ. An areawide examination of nursing home use, misuse, and non-use. Am J. Public Health 1971;61:1146.

3. Williams FT, Hill JG, Fairbank ME, Knox KG. Appropriate placement of the chronically ill and aged: a successful approach by evaluation. JAMA 1973;226:1332.

4. U.S. Department of Health, Education and Welfare. Problems in providing proper care to Medicaid and Medicare patients in skilled nursing facilities. Washington, D.C.: U.S. Government Printing Office, 1971. (Report No. B-164031(3).)

5. Zimmer JG. Characteristics of patients and care provided in health-related and skilled nursing facilities. Med Care 1975;12:992.

6. Allison-Cooke S, Thornberry H. Factors affecting nursing home medical review: implications for program and facility planning. Med Care 1977;15:494.

7. Connelly K, Cohen PK, Walsh DC. Periodic Medical Review: assessing the quality and appropriateness of care in skilled nursing facilities. N Engl J Med 1977;296:878.

8. Rowland KF. Environmental events predicting death for the elderly. Psychol Bull 1977;84:349.

9. Beverly EV. Nursing homes: matching the facility to the patients' needs. Geriatrics 1976;31:100.

10. Katz S, Amasa BF, et al. Studies of illness in the aged: the index of ADL: a standardized measure of biological and psychosocial function. JAMA 1963; 185:914.

11. Schulz R. Effects of control and predictability and the physical and psychological well-being of the institutionalized aged. J Pers Soc Psychol 1976;33:563.

12. Kahana E, Coe RM. Alternatives in long-term care. In: Sherwood S, ed. Long-term care: a handbook for researchers, planners, and providers. New York: Spectrum Publications, Inc, 1975:511-72.

13. Seligman ME. Helplessness: on depression, development, and death. San Francisco: Will Freeman and Company, 1975.

14. U.S. Department of Health, Education and Welfare. Medicaid: medical review in skilled nursing homes and mental hospitals: guidelines. Washington, D.C.: Medical Services Administration, 1972.

15. Gaynes NL. A logic to long-term care. Gerontologist 1973;13:277.

16. Yokie JA. A comparison: long-term care versus acute care. Nursing Homes 1977;27:18.

17. Kosberg JI. Making institutions accountable: research and policy issues. Gerontologist 1974; 14:510.

18. Ruchlin HS. A new strategy for regulating long-term care facilities. J Health Polit Policy Law 1977;2:190.

19. Densen PM, Jones EW. The patient classification for long-term care developed by four research groups in the United States. Med Care 1976;14(supplement):126.

20. MacDonnell JA. Canadian experience with patient care classification. Med Care 1976;14(supplement):134.

21. Anderson NN, Stone LB. Nursing homes: research and public policy. Gerontologist 1969;9:214.

22. Sherwood S. Long-term care: issues, perspectives, and directions. In: Sherwood S, ed. Long-term care: a handbook for researchers, planners, and providers. New York: Spectrum Publications, Inc, 1975:3–79.

23. Kosberg KI. Methods for community surveillance of geriatric institutions. Public Health Rep 1975;90:145.

24. Jaffee LL. The effective limits of the administrative process: a reevaluation. Harvard Law Review 1954;67:1105.

25. Bernstein MJ. Regulating business by independent commission. Princeton: Princeton University Press, 1955.

26. Noll RG. The consequences of public utility regualtion of hospitals. In: Institute of Medicine. Controls on health care. Washington D.C.: National Academy of Sciences, 1975.

27. Fainsod M. Some reflections on the nature of the regulatory process. In: U.S. Senate. Separation of powers and the independent agencies: cases and selected readings. Washington, D.C.: U.S. Government Printing Office, 1970. (Document No. 91-49.)

28. Mendelson MA. Tender-loving greed: how the incredibly lucrative nursing home industry is exploiting America's old people and defrauding us all. New York: Alfred A. Knopf, Inc, 1974.

29. Townsend C. Old age: the last segregation. New York: Grossman Publishers, 1971.

30. U.S. Senate, Subcommittee on Long-Term Care. Nursing home care in the United States: failure in public policy: introductory report. Washington D.C.: U.S. Government Printing Office, 1974.

31. Pennsylvania Office of State Planning and Development. Pennsylvania population projection series: population and labor force. Harrisburg, 1973.

32. Pennsylvania Department of Public Welfare, Office for Aging. Serving older persons in Pennsylvania: a summary report and proposals for the state plan on services for older people. Harrisburg, 1977.

33. Pennsylvania Department of Commerce, Bureau of Statistics, Research, and Planning. Pennsylvania statistical abstract. Harrisburg, 1977.

34. Pennsylvania Department of Commerce, Bureau of Statistics, Research, and Planning. Pennsylvania statistical abstract. Harrisburg, 1978.

35. Pennsylvania Office of the Budget, Division of Program Planning and Evaluation. Nursing homes in Pennsylvania: a patient profile. Harrisburg, 1975.

36. Pennsylvania Department of Public Welfare, Bureau for the Aging. The elderly in Pennsylvania: data packet no. 1: demographic information. Harrisburg, 1974.

37. Davis JJ, Gibbin MJ. An areawide examination of nursing home use, misuse, and nonuse. Am J Public Health 1971;61:1146.

38. Roemer M. Bed supply and hospital utilization: a natural experiment. Hospitals 1961;35:36.

39. Feldstein MS. Economic analysis for health service efficiency: econometric studies of the British national health service. Chicago: Markham Publishing Company, 1968.

40. Anderson JG. Demographic factors affecting health service utilization: a causal model. Med Care 1973;11:104.

41. Willemain TR, Farber ME. Nursing homes and the Roemer-Feldstein hypothesis. Med Care 1976;14:880.

MEDICAL CARE
January 1982 Vol. XX, No. 1

Alcoholism, Morbidity and Care-Seeking

The Inpatient and Ambulatory Service Utilization and Associated Illness Experience of Alcoholics and Matched Controls in a Health Maintenance Organization

SANDRA L. PUTNAM, M.SC.

The issue of whether outpatient treatment of alcoholism is cost-advantageous, in the long run, for health maintenance organizations (HMOs) depends in part on whether alcoholics represent a relatively heavy burden on the health care delivery system. To clarify this issue, the paper asks whether alcoholics utilize HMO services at higher rates and in different ways than do other HMO members, and whether alcoholics experience more illnesses and injuries associated with their service utilization. The study subjects were alcoholics, identified during one year as new clients of an HMO's counseling department, which houses an alcoholism treatment component. Study subjects were matched with controls on the basis of sex, age, date of HMO enrollment, type of membership and family size. HMO medical records provided data on service utilization and associated morbidities during the three-year study period. Alcoholics were found to have utilized about 50 per cent more of all HMO services studied than did their matched controls. Alcoholics were especially high utilizers of more expensive, inpatient services. Psychosocial problems and problems classified as accidents, poisonings and violence were much more likely to underlie alcoholics' hospitalizations, and, to a lesser extent, their outpatient non-counseling utilization, than was the case for controls. Some tendency was noted for more chronic illnesses to be associated with alcoholics' service utilization, and more acute illnesses with controls' utilization.

ALCOHOLISM is frequently characterized as an illness, the economic, medical and social costs of which reach beyond the alcoholic individual and his or her social network to affect the larger community, through alcohol-related motor vehicle and other accidents, crimes, sickness absenteeism, lost earnings and lowered productivity in the work place, and high health care costs. Some researchers have tried to estimate the actual costs of alcoholism and the benefits obtained from its treatment.[1–3] However, very little has been written to document how alcoholics incur high health care costs, other than through hospitalizations and sickness and disability benefits, and how they differ from other consumers of medical care services.

The purpose of this paper is to address the cost issue indirectly by examining some of the results of a medical care utilization research project. The study was conducted at Rhode Island Group Health

---

From the Occupational Alcoholism Project, Rhode Island Group Health Association, 530 North Main Street, Providence, Rhode Island 02904.

This paper was presented in modified form at the 1980 Group Health Institute in Boston on May 6, 1980, where it was entitled, "A Comparison of the Utilization of HMO Services by Alcoholics and Matched Controls."

Supported by grant 5H84-AA-02210 from the National Institute on Alcohol Abuse and Alcoholism.

0025-7079/82/0100/0097/$01.75 © J. B. Lippincott Co.

Association, a health maintenance organization (HMO) in Providence, Rhode Island. It is a partial replication of the Genesee Valley Group Health Association's utilization project, reported by Wersinger and his colleagues.[4]

The principal research question is whether alcoholic HMO members utilize more medical care services for more and different types of illnesses and injuries than do matched controls. Such an inquiry represents one way of determining whether alcoholics place a relatively heavy burden on a closed health-care delivery system, such as an HMO. This would be expected in light of their well-documented tendency to experience high rates of certain illnesses and injuries secondary to, or associated with, alcoholism.[5] For example, among the chronic, degenerative diseases found to be positively associated with alcoholism in one non-hospital-based morbidity study were cirrhosis of the liver and hypertension, as well as peptic ulcer, asthma, diabetes, gout, neuritis, cerebrovascular disease and heart disease.[6] Other alcohol-related morbidities identified in studies using hospitalized subjects include, besides most of the above, other liver diseases, pancreatitis, tuberculosis, certain cancers, gastrointestinal problems, such as gastritis, seizure disorders, and accidents and traumas.[7–12] Many of the same illnesses are directly associated with alcoholics' mortality rates. These have been consistently found to be two or three times higher than expected rates, with excess deaths of alcoholics in most, if not all, major disease categories.[13–15]

If ambulatory, inpatient, and other HMO service utilization studied is significantly higher among alcoholic study subjects than among controls, then the need for third-party insurance coverage of alcoholism treatment and the desirability of establishing alcoholism treatment programs in HMOs may be more readily demonstrated. If differences in the nature and prevalence of morbidities associated with service utilization are found, then elucidating the physical and emotional problems of which alcoholics are at very high risk may constitute a practical means of signaling possible alcoholism among patients. The illnesses and injuries, which are viewed here as "signals," often may serve as smoke screens, obscuring the primary problem of alcoholism from medical care providers, who are already treating many alcoholics under other diagnoses.[16]

Identifying these signals of possible alcoholism may increase the chances of its early detection and timely intervention in the disease process. The chances may also be enhanced of treating the secondary or associated illnesses effectively, whether they are exacerbated by alcohol or are directly caused by the hidden, untreated alcoholism. While alcohol is clearly a dangerous drug, nevertheless, the behaviors associated with alcoholism, such as other drug use, heavy cigarette smoking, poor diet, social isolation and other lifestyle factors, may all contribute to explaining the relationship of alcoholism with excess morbidity and mortality.

Background of the Study

The HMO Setting

Rhode Island Group Health Association (RIGHA) is the only federally qualified HMO in Rhode Island. It was incorporated in 1968 and instituted services in 1971. Enrollment has increased from 1,200 in 1971 to 12,000 in 1974 to 30,000 in January 1979. The main HMO health center is staffed by 300 employees, including 27 salaried physicians. It houses, besides business and administration offices, a full laboratory and x-ray facility, a pharmacy, a dental department, vision services, and a multi-specialty area, where clinics are held in dermatology, allergy, orthopedics and other specialties. Departments of pediatrics, obstetrics-gynecology, internal medicine, surgery and counseling are also located at the health center, as is the urgent visit unit, providing 24-hour telephone coverage for emergency care, and the in-house occupational alcoholism project.

HMO coverage costs, on average, about $20 more a month than the experience-rated Blue Cross-Blue Shield plan. Most members join through employer groups, for which there are no pre-existing conditional requirements and of which there are about 900 offering RIGHA coverage. This coverage includes 100 per cent of all ambulatory care delivered by HMO providers with a few exceptions, such as most dental services and counseling services, for which fees are charged to members. These services are excluded from this study, since their inclusion would change the situation that otherwise pertains, whereby the effects of income differentials are controlled. HMO coverage also includes 100 per cent of semi-private hospitalization, together with out-of-state hospital coverage with health plan authorization. Detoxification and rehabilitation of alcoholics at area inpatient and residential facilities are covered under this benefit. Referrals by RIGHA providers to non-plan specialists are also fully covered. The extent of unauthorized, out-of-plan utilization of health care services by RIGHA members is unknown, but it is thought to be minimal.

Characteristics of the HMO over the Study Period

Over the study period, which consists basically of three calendar years—1976, 1977 and 1978— the health center was relocated to a larger and more centralized site, and a satellite clinic was opened 12 miles away. The number of non-primary physicians has increased through the addition of several part-time medical specialists, including a neurosurgeon, a urologist, and an ear, nose and throat specialist, who hold weekly in-house clinics and are paid by capitation. This trend has been accompanied by a decrease in the number of very expensive external referrals, that is, those made to non-HMO providers, usually specialists. Such external referrals have decreased by half over the study period—from 8.1 per 1000 enrollees in 1976 to 3.8 per 1000 in 1978. The number of primary physicians in ob-gyn and medicine has remained about the same, ranging from 16 to 19, over the study period, with a growing tendency to use more paraprofessionals to deliver primary medical care services. Indeed, the number of ancillary personnel—physicians' assistants, nurses, nurse practitioners and health assistants—has doubled, from 13 in 1976 to 25 in 1978. Finally, ambulatory utilization during the study period increased from about 3.2 visits per person per year in 1976 to 4.3 visits per person per year in 1978, with hospitalization rates decreasing from 402 hospital bed days per 1000 members in 1976 to 355 bed days per 1000 in 1978.

The Occupational Alcoholism Project

Fairly unique additions to the services available in this HMO setting are those provided through the outreach and treatment components of the occupational alcoholism project. The project has been funded by a grant from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) since September 1975. It has received accreditation from the Joint Commission on the Accreditation of Hospitals since 1977. It uses the employee assistance approach to the identification, confrontation and referral, by supervisors, of employees with job performance problems, which frequently mask alcohol problems. The outreach component of the occupational alcoholism project is the subject of an earlier research report.[17]

The project uses the HMO's counseling department as the principal treatment resource. Complete intake, evaluation and outpatient treatment services are offered by 12 therapists to HMO members and to employees of organizations with employee assistance programs for alcohol as well as for psychiatric problems. The alcoholism treatment program, then, is integrated into the mental health department of the HMO. It is associated with the outreach component of the occupational alcoholism project. It operates within the closed health care delivery system of the HMO, insuring information-sharing through medical records and easy access to medical care providers, with a family-oriented approach. All of these conditions are thought to increase the likelihood of early detection and intervention in alcoholism in this setting.

The counseling department has grown from serving 656 new clients or "intakes" in 1976 to 714 in 1978. An average of 63 new clients a month came to counseling during the study period. Thirty per cent of these intakes, on average, were diagnosed as having the primary problem of alcoholism. Another 27 per cent of intakes were spouses or other family members of alcoholics. Such persons are classified as having "alcohol-related" problems. Of the alcohol-involved intakes (including alcoholics and family members), about one quarter were referred directly by supervisors at their place of employment. One third were referred through a health care provider, and 37 per cent were self-referred or were referred by family or friends.

The presence of the occupational alcoholism project in the HMO made it relatively easy to identify a group of alcoholics, who were HMO members and for whom, therefore, complete medical and counseling records were available. Wersinger and his colleagues[4] lacked such ease of access to an alcoholic sample. In addition, the diversity of referral sources of alcoholics to the counseling department may insure that less severely troubled alcoholics were included in the study group than are often caught in samples of alcoholics. One possible indication that relatively early detection of alcohol problems is occurring at RIGHA is the fact that, on average, study subjects are younger (mean age is 37.3 years) than either the inpatient alcoholics (mean age is 45.8 years) or the confirmed alcoholics (with a mean age of 41.1 years) who served as study subjects in the Genesee Valley Group Health Association project.[4] Table 1 shows the age and sex distribution of alcoholics in the present study.

Relevant Literature

Studies of Alcoholics' HMO Service Utilization

Whether identified alcoholics utilize prepaid medical care services at higher rates and in different ways than do other HMO members has been the topic of two previous studies in the United States. The first, based on Boyajy's published reports,[18,19] concerns a four-year research and demonstration project undertaken by the Group Health Association of America and funded by a grant from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Four geographically distant health plans provided data on a variety of subjects, including the utilization behavior of identified alcoholics. Although the failure to use matched control groups flawed the research, as the author notes, still, the findings were as expected.[19] The alcoholics consumed services at a significantly higher rate than other group members (ref. 19, page 419). The overall utilization rate of alcoholics at one plan, for instance, was 22 per cent higher than that of the general health plan population, controlling for age. Further, the patterns of use were found to be different for the alcoholics, who showed a predominance of unscheduled and other relatively expensive visits, including high utilization of physician services and emergency visits (ref. 18, pages iii–iv).

Boyajy also hypothesizes that alcoholics will tend to have more illnesses with a "high emotional component" than other HMO members. In support of this contention, he lists the ten most frequent diagnoses associated with alcoholics' utilization of services. One of these is anxiety, and

TABLE 1. Comparison of the Age–Sex Distributions of Study (and Control) Subjects with that of the Total RIGHA Membership during 1979

Age Groups	Age–Sex Distributions								
	Study (Control) Group*						Overall RIGHA Membership†		
	Males		Females		Total		Males	Females	Total
	#	%	#	%	#	%	%	%	%
15–24	4	5.3	5	16.7	9	8.5	25.0	27.6	26.3
25–44	51	67.1	20	66.6	71	67.0	43.7	42.4	43.1
45–64	20	26.3	5	16.7	25	23.6	27.0	26.6	26.8
65+	1	1.3	0	0	1	0.9	4.3	3.4	3.8
Total	76	100.0	30	100.0	106	100.0	100.0	100.0	100.0
	$\bar{X}$ = 38.8		$\bar{X}$ = 33.5		$\bar{X}$ = 37.3		Median = 27.8	Median = 27.2	Median = 27.6

* Because controls were matched with alcoholics by age and sex, among other characteristics, the age–sex distributions for the two groups are identical.

† These percentages of RIGHA members in age–sex groups are based on the averages of figures for two points in the study period, May 1, 1976 and December 31, 1978.

some others are illnesses that are thought to be stress-related. In descending order of frequency, these diagnoses are alcoholism, hypertension, upper respiratory tract infection, influenza, trauma, alcoholic gastritis, gout, anxiety reaction, contusion, cirrhosis, and allergic reactions. Ulcers and drug reactions join the list in the six-, twelve- and eighteen-month follow-up data.[18] While these results are neither discussed in detail by Boyajy nor compared with a similar list of morbidities for a nonalcoholic or control group, they provide a basis for comparison with this study's results. They also give rise to one of the hypotheses tested here.

Related to Boyajy's study is the Genesee Valley Group Health Association's utilization project alluded to earlier.[4] In this project, three study groups of alcoholics were drawn from the HMO population, with some difficulty, since the HMO lacked both a psychiatric or counseling department and an alcoholism treatment component. The subjects consisted of a group of 10 inpatient alcoholics, who were identified through hospital claims forms; 59 confirmed alcoholics, who were diagnosed as such by HMO providers; and 98 suspected or high-risk alcoholics, who were identified by answers to questionnaires. The subjects were matched on a number of sociodemographic dimensions to individuals in three control groups. Utilization data were provided by hospital claims forms and ambulatory encounter forms.

The authors found, as hypothesized, that the alcoholics (inpatients and confirmed alcoholics grouped together) experienced higher rates of utilization of both inpatient and ambulatory services compared with controls. The largest differences in utilization were found for inpatient hospitalizations. This is not unexpected, since the inpatient group was selected in such a way that a bias in favor of higher rates was ineluctably introduced. Still, confirmed alcoholics utilized eight times as many inpatient services as their controls. Most of the differences between alcoholics and controls were attributed to psychiatric admissions, which were 100 times higher for alcoholics than for controls, and, to a lesser extent, to medical admissions, which were four times higher for alcoholics. No differences in the rates were observed for surgical admissions.[4]

As for ambulatory utilization, measured in terms of average number of encounters per person-year over a four-year period, the alcoholics had 23 per cent higher rates than the control groups. Much of the difference in the encounter rates was due to a 50 per cent higher illness rate for alcoholic study subjects.[4] The fact that medical hospital admissions were relatively high for alcoholics and that they had higher ambulatory encounter rates for the treatment of medical problems led Wersinger and his co-authors to suggest that alcoholics are higher utilizers of services, because they are more likely to be ill than controls.

This explanation of alcoholics' relatively high utilization is reinforced by the general finding for nonalcoholic groups that need for care or poor health is the strongest predictor of physician utilization in the fee-for-service sector.[20] More specifically, it is suggested by Boyajy's finding of a relationship among degree of alcohol involvement, level of morbidity, and utilization rates. He found that the most severely troubled alcoholics, whose consumption levels were highest and whose duration of heavy drinking was longest, seemed to be higher utilizers of services and were at the greatest risk for a number of morbidities.[18]

In addition to their greater need for care, another explanation is possible for alcoholics' higher utilization of services, when age, sex, and other sociodemographic factors are controlled. It is that they report, rather than necessarily experience, more illnesses, that they are more likely than controls to seek treatment and to be compliant in following up on it, and that they, therefore, enter the reporting system more often. This notion is difficult to test, since it is founded on the concept of illness behavior, rather than illness itself, as the independent variable for explaining utilization. However, because alcoholism is said to be an illness, the denial of which is a key symptom, and because alcoholics are ". . . notoriously non-compliant,"[12] they might be more apt to avoid medical treatment than to seek it out in circumstances that could lead to their identification as chemically dependent by medical care providers. This issue is important and requires further study.

Hypotheses

Based on the literature, the hypotheses explored in this study are:

1. The alcoholic study subjects will be found to be higher utilizers than controls of all types of services studied, particularly of inpatient hospital services, even after eliminating those for alcoholism detoxification and rehabilitation, and
2. that the study subjects will be found to experience (or report) more illnesses, both physical and emotional, and more injuries than controls.

The following three criteria are used to specify and operationalize the second hypothesis:

a. that alcoholic study subjects will report more problems or morbidities than controls;

b. that alcoholic study subjects will have more problems with a high emotional component, including primarily what were coded in this study as psychosocial problems; and

c. that study subjects will tend to report more serious or chronic problems than controls.

Data and Methods

Sample Selection

The experimental group for this study was drawn from the population of new counseling clients reported as primary case members with the primary problem of alcohol on NIAAA Initial Contact Forms, between January 1, 1977 and March 31, 1978. These forms are filled out by therapists for all counseling clients on intake or first counseling encounter, both to satisfy the NIAAA grant requirements and for internal recordkeeping purposes. A further criterion for eligibility in the study group was continued membership in the HMO, with no more than a three-month lapse, during the entire 39-month study period, from January 1, 1976 to March 31, 1979. This insured that all study subjects possessed medical records, which served as the principal data source, and that utilization data were available on every subject for the entire study period and for at least one year before and one year after alcoholics' counseling intake date or its equivalent for controls.

A total of 294 alcoholics were identified when entering counseling for the first time during the 15-month pivotal period—a total of 242 during calendar 1977 and another 52 during the first three months of 1978. Of these potential study subjects, about half were eliminated, because they lacked appropriate insurance coverage, having either no insurance or other than HMO coverage. About one third (51) of the 157 RIGHA members were ineligible for inclusion, because their HMO coverage was not sufficiently extensive. The remaining two thirds of the alcoholics who were HMO members made up the study group. Parenthetically, the spouses of the study subjects and of the controls were also included in the study. The same data were collected on their service utilization as was on their spouses' utilization. For purposes of this paper, however, data on the 76 study spouses and 76 control spouses are omitted.

Controls were selected from lists of current RIGHA members, as of March 31, 1979. They were matched to study subjects on basically the same dimensions used for matching in the Genesee Valley study.[4] These are sex, date of birth (age), date of effective HMO membership, type of HMO membership (individual or family, which often indirectly indicates marital status for adults), and number of dependent children. The task of matching, although done by hand rather than by computer, was facilitated by the fact that RIGHA identification numbers are assigned, in part, on the basis of date of enrollment, and microfiche membership files are organized by identification number. Originally, several controls were matched to every study subject. The controls who were kept in the study, after matching on sex and type of membership, were matched, as closely as possible, with alcoholic study subjects by date of birth, then by date of effective HMO membership, and finally, as the least critical matching variable, on number of dependent children. One hundred and six study subjects and the same number of matched controls were included in the study.

A possible source of bias in this study is that some potential controls, as well as some potential study subjects, were HMO members who were ineligible for inclusion in the study, because they joined RIGHA after January 1, 1976 or cancelled before March 31, 1979. Cancellations are due to job changes, with consequent loss of HMO group coverage; to geographical mobility; to divorce; in rare cases, to death; and to voluntary changes of coverage due to dissatisfaction with the health plan. Alcoholics may be more likely to experience loss of HMO coverage than other plan members. Eliminating HMO "drop-outs" from the study introduces a bias in favor of stability of life circumstances and satisfaction with the plan. In the case of alcoholics, it may favor the inclusion of those who are less severely troubled or less severely impaired. Since this would lead to greater similarity between alcoholics and controls, less difference in their utilization would be expected. Therefore, any substantial differences that are observed in utilization of HMO services, measured by encounters, and in associated problems or morbidities between experimental and control groups will gain added weight.

Characteristics of Study Subjects and Controls

The age distribution of the 106 study and matched control subjects is compared with that of the overall HMO population during the study period in Table 1. It indicates that the sample underrepresents young (15–24) and elderly (65+) RIGHA members, while subjects in the early working ages (25–44) are overrepresented, compared with the plan membership as a whole. The sex distribution is different from that of the HMO overall, where females are about equally represented. For the study and control groups, 29 per cent are female. In the sample, female subjects are slightly younger than males, with a mean age of 33.5 years, compared with a mean age of 38.8 years for males. The ratio of members to subscribers is also higher in the sample than in the plan—3.2 compared with 2.3. A subscriber is an HMO member who holds the membership contract; members include all subscribers plus the eligible dependents of subscribers. Subjects have an average of 2.1 dependents and 1.4 dependent children per subscriber. The sample is thus skewed toward family memberships (married persons). Seventy per cent of subjects had family memberships and 30 per cent had individual memberships, compared with 56 per cent family memberships and 44 per cent individual memberships (excluding Plan 65+ members) for the total RIGHA population, as of January 31, 1979. The fact, mentioned earlier, that the sample alcoholics are younger, on the average, than many other groups of alcoholics who have been studied, might be reflected in comparatively low utilization rates for both alcoholics and controls, since age is associated with illness.[21,22]

Besides the group's relative "youth" and the high representation of women (one third), compared with other samples with less than 10 per cent women, another fairly unique aspect of the study group is that it is a non-hospitalized sample. This is unusual in alcoholism morbidity studies.[10] Secondly, that it includes both acceptors of HMO alcoholism treatment and non-acceptors, those who dropped out of treatment, is also exceptional. Other studies have used, as subjects, voluntary participants in long-term alcoholism treatment, who differ in unknown ways from untreated alcoholics.[8] This study group may thus represent an infrequently sampled population of alcoholics.

Measures of Utilization

The "encounter" is the main unit of analysis according to the coding system developed for use with the medical records. An encounter is defined for a subject as a face-to-face contact with an HMO provider (physician, physician's assistant, nurse practitioner, etc.) or with a non-plan or external provider. Encounters include visits for laboratory tests and x-rays, and hospitalizations. Ambulatory encounters encompass health maintenance unit visits for initial and annual physical examinations, as well as most primary care and specialty visits. Dentistry and optometry encounters were excluded, as were prenatal care clinic visits and periodic visits with nurses for the sole purpose of blood pressure checks. Reports of some such encounters are kept separately from the HMO medical records. In other cases, the reporting of such encounters in the medical records was found to be inconsistent and incomplete. Physical therapy encounters were coded when external referrals were made. HMO-based physical therapy encounters were coded if they were reported in the medical records, which is thought to be fairly often. The extent of uneven coverage of this category is unknown. No well-child and pediatrics services were used by the subjects. Counseling encounters were coded, using counseling department records rather than medical records, but they were omitted from this analysis, which is confined to medical care utilization behavior.

Due to the definition of encounters chosen, telephone calls and letters were not included in the coding, unlike the situation in Boyajy's study. Hospital visits by doctors and home visits by nurses, which occur very infrequently, were not coded, because they are not routinely reported in the medical records. All other ambulatory encounters, all external referrals, all laboratory and x-ray procedures, and all hospitalizations were coded for all subjects. Each separate encounter was coded by type or department—medical, surgical, orthopedic, etc.—by date, by associated morbidities (called "problems"), by procedures reported done, and by the type of provider seen.*

Both numbers and proportions of encounters and procedures are used as measures of utilization in this study. "Procedures" are defined to include preventive procedures (such as certain physical examinations), follow-up encounters after initial appointments, and diagnostic procedures, such as treadmill stress tests, EKGs, nerve conduction studies, etc. Also included, particularly in the inpatient utilization category, are surgical procedures. These were coded using a somewhat revised and condensed version of the Blue Shield Schedule of Relative Values.[23] Therapeutic procedures, particularly certain drug therapies, were also coded.

The overall ratio of procedures to encounters for alcoholics and controls is 0.93, as shown in Table 2. Procedures are thus fairly good proxy measures for encounters, especially for alcoholics whose overall ratio is 0.96. Using procedures to measure utilization overestimates external referrals by 8 per cent, and somewhat underestimates outpatient encounters, especially for controls. The greatest discrepancy between the two utilization measures is in the inpatient category, where, overall, 26 per cent more procedures were coded for the average encounter, with controls' utilization being more distorted by measuring procedures than alcoholics' utilization. Encounters are therefore the measure of choice here.

"Problems" are defined as including symptoms reported by the patient when the provider failed to or was unable to diagnose the illness presented, as well as providers' diagnoses and assessments. Problem codes are based on the three-digit series of codes in the eighth edition of the International Classification of Diseases (ICD).[24] The only major revisions of these codes were the substitution of other psychosocial codes for mental disorder codes and some coding changes in the category of "Symptoms and Ill-Defined Conditions."

TABLE 2. The Ratios of Procedures to Encounters by the Three Main Service Utilization Categories for Alcoholics and Controls

Service Utilization Categories	Ratios of Procedures to Encounters		
	Alcoholics	Controls	Overall
External referrals	1.08	1.08	1.08
Inpatient encounters	1.22	1.36	1.26
Outpatient non-counseling encounters	0.94	0.85	0.90
Total	0.96	0.88	0.93

---

* The way in which information was coded on hospitalizations for any encounter or admission is illustrated by the case of a hypothetical encounter, which was a medical emergency, on 8/20/77, for the problems of facial lacerations, cerebral concussion and adverse reaction to an analgesic drug. Procedures included sutures and tomography. The length of stay was 5 days. An external referral was coded, for example, as occurring at a vascular laboratory, on 11/15/78, for the problems of lower leg trauma and thrombophlebitis, the procedure being venous evaluation. Information collected for coding ambulatory encounters included the fact that the patient, for example, on 5/4/77, was seen in the HMO surgery department for a dog bite on the ankle, which required sutures and a tetanus shot, which were provided by a physician. Another example is on 1/30/76, a patient was seen in internal medicine for sleep problems, epigastric pain, and anxiety, and was given a physical examination and a prescription for flurazepam by a physician (Certain selected drugs, particularly CNS depressants, were coded in this study, when they were reported as prescribed.)

Since this is a descriptive study, group differences for the three main service utilization categories—external referrals, hospitalizations, and ambulatory encounters—are examined, using percentage distributions and indexes of dissimilarity (ID) to summarize distribution differences.† Relative risk, defined as the ratio of the alcoholic group's score or rate to that of the control group,[4] is also employed to portray group differences. For the most important comparisons, Student's t-tests for matched or dependent samples are calculated as tests of significance.[25]‡

Results and Discussion

Utilization of HMO Services by Alcoholics

The first hypothesis of this study was supported, as shown in Table 3. Overall, the alcoholic study subjects utilized one-and-a-half times more services, measured by encounters, than did controls. The magnitude of the difference between alcoholics and controls (45 per cent) is greater than was found by Boyajy (22 per cent).

As expected, the difference in service utilization was greatest for hospitalizations. Alcoholics had 143 per cent more inpatient encounters (separate hospitalization episodes) than did controls. This difference is significant at the 0.001 level (t = 3.82, df = 105). Eliminating hospitalizations for detoxification from inpatient encounters reduces the difference between the two groups only slightly, with study subjects still utilizing about twice as many inpatient services (the relative risk is 2.02). The differences remain significant (t = 4.29, df = 105).

According to the relative risks in Table 3, alcoholics made 43 per cent more ambulatory non-counseling encounters, compared with controls. This difference is significant at about the 0.001 level (t = 3.42; df = 105). Study subjects utilized 45 per cent more laboratory and x-ray services§ and 28 per cent more external referrals. External referrals comprise the utilization category that least differentiates alcoholics and controls. The group differences here do not reach significance (t = 0.933; df = 105).

Not only did alcoholics use more medical services after seeking care at the HMO, more alcoholics than controls sought such medical care in the first place. More alcoholics were users of HMO services. More controls failed to use such services. This is shown in Table 4, where not numbers of encounters, but numbers of subjects making encounters, are compared. While alcoholic study subjects and controls were similar in their nearly universal utilization of outpatient services and laboratory and x-ray procedures, about half as many controls as alcoholics were hospitalized (52 alcoholics compared with 29 controls) and about one third as many controls were referred to non-plan providers (36 per cent of controls, compared with 53 per cent of alcoholics).

Table 4 also gives the encounters per user in each utilization category over the three-year study period. Alcoholic users had more encounters, on average, than controls in three of the four service utiliza-

---

† The index of dissimilarity is interpreted as the percentage of cases that would hypothetically need to be redistributed to make two percentage distributions equal to one another. The formula is $\frac{1}{2}\sum|P_1 - P_2|$, where P = percentage distribution.

‡ For these significance tests, each pair of individuals—each alcoholic and his/her matched control—is treated as a single case. Direct pair-by-pair comparisons of encounters are made by obtaining a difference score for each pair and testing the null hypothesis that the two populations do not differ and therefore that alcoholism has no effect on utilization and morbidity. A single-sample test of the hypothesis that the standard deviation of the differences equals zero is used. T-tests are calculated, and the null hypothesis is rejected at the .05 level of significance or below.[25]

§ When a subject was referred to the HMO laboratory, a number of tests were usually ordered done at once, so that a count of procedures rather than encounters, gives a more accurate picture of the utilization of such services.

TABLE 3. Numbers of Encounters Made by Alcoholics and Controls by Service Utilization Category, Including Laboratory Tests and X-Rays, External Referrals, Hospitalizations and Ambulatory Encounters

Service Utilization Categories	Numbers of Encounters		
	Alcoholics	Controls	Relative Risk
Laboratory tests and x-rays*	1780†	1230†	1.45
External referrals	143	112	1.28
Inpatient encounters (hospitalizations)	107(89)‡	44(44)‡	2.43(2.02)‡
Ambulatory non-counseling encounters§	1695	1184	1.43
Total	3725	2570	1.45

* Laboratory tests and x-rays are counted by procedures, rather than by encounters, because a number of such tests were usually done during one encounter. Procedures are therefore a more accurate measure of utilization in this service utilization category than encounters.

† These laboratory tests and x-ray procedures exclude a few codes used for drugs administered during these procedures, once for an alcoholic and six times for four controls.

‡ The figures in parentheses are the results, when alcoholism detoxification and rehabilitation hospitalizations are excluded.

§ Ninety-one ambulatory counseling encounters made by alcoholics were excluded from the original data, as they had been mistakenly included. The results shown omit these encounters. Thirteen such encounters for controls were similarly omitted. Their inclusion raised the relative risk slightly to about 49 per cent for ambulatory encounters. It also changed the ratios of outpatient procedures to outpatient encounters for alcoholics from 0.94 to 0.95, while the ratio for controls remained unchanged at 0.85.

tion categories. These included an average of 2 hospitalizations, compared with 1.5 for control users; 16 outpatient encounters, compared with 11.5 for controls; and 17 laboratory and x-ray procedures, relative to 13 for controls. Unexpectedly, however, more external referrals were made for and/or accepted by control users than was the case for alcoholics. An average of 3 external referrals was recorded for control users, while 2.5 was the alcoholics' average. It may be that alcoholics, who are sometimes regarded as "bad," i.e., difficult and undesirable patients, are not being as readily referred to non-plan specialists by HMO providers, despite their higher morbidity. Alternatively, it may be that as many or more such referrals are made, but they are not as readily accepted by alcoholics as by controls. It is also possible that although more alcoholics were referred to external providers than controls, the latter users,

TABLE 4. Numbers of Subjects Utilizing Services (Users) and Average Encounters Per User by Service Utilization Categories for Alcoholic Study Subjects and Matched Controls

Service Utilization Categories	Alcoholics			Controls		
	No. Users	% Potential Users (106)	Encounters Per User	No. Users	% Potential Users (106)	Encounters Per User
Laboratory and x-ray procedures	105	99.1	17	95	89.6	13
External referrals	56	52.8	2.5	38	35.8	3
Inpatient encounters	52	49.1	2	29	27.4	1.5
Outpatient non-counseling encounters	105	99.1	16	103	97.2	11.5

TABLE 5. Numbers and Percentages of "Problems" or Morbidities Reported During Encounters, by Service Utilization Categories, for Alcoholic Study Subjects and Controls

Utilization Categories	Total Occurrences of Problems or Morbidities*				Relative Risk
	Alcoholics		Controls		
	No.	%	No.	%	
External referrals	200	6.6	180	9.5	1.11
Inpatient encounters	258	8.6	88	4.7	2.93
Outpatient non-counseling encounters	2558	84.8	1622	85.8	1.58
Total reported problems associated with service utilization	3016	100.0	1890	100.0	1.60

* Total occurrences of problems are counts of the numbers of separate problem codes, summed over all problem categories, for all subjects and for distinct utilization categories. Therefore, for example, alcoholics' 2558 outpatient problems included 84 occurrences of diabetes, 64 occurrences of obesity, 129 occurrences of hypertension, etc.

while fewer in number, may have been more severely ill or injured, requiring more external services. This issue deserves further study.

Parenthetically, the types of procedures that were most frequently utilized differed both by utilization category and by group. The most striking difference between alcoholics and controls was in the marked prevalence of sterilization procedures incurred by the study group. Four of the 30 alcoholic women in the study sample were hospitalized for tubal ligations during the three-year study period, compared with no controls. Similarly, twice as many vasectomies (6) were performed, on an outpatient basis, on alcoholic men as on controls (3). Because this is essentially a pilot study, no conclusions can be drawn here. However, the questions of whether alcoholic women have higher encounter rates for obstetrical and gynecological problems and whether both male and female alcoholics have more sex-related problems than their respective controls are worthy of further study.

Illnesses and Injuries Associated with Alcoholism

The second hypothesis, that alcoholics will be more likely to experience (or report) illness than controls, was supported, although more intensive data analysis is needed to further specify the relationship between levels and types of morbidity and utilization behavior.

As Table 5 indicates, alcoholic study subjects reported 60 per cent more problems associated with their overall service utilization than did controls. The largest difference between the two groups was in hospitalizations, where nearly three times as many problems were recorded for alcoholics as for controls. For each inpatient encounter, 2.4 problems, on average, were recorded for alcoholics, compared with an average of 2 problems per encounter for controls. This suggests that when study subjects were hospitalized, they were diagnosed as having more associated problems than were hospitalized controls. For

---

A number of the results reported here are slightly different from those reported in an earlier version of this paper, entitled "A Comparison of the Utilization of HMO Services by Alcoholics and Matched Controls" by Sandra L. Putnam, presented at the 1980 Group Health Institute, May 6, 1980, in Boston, Massachusetts. This is because the original results contained data on some ambulatory counseling encounters, which somewhat inflated the psychosocial problem counts. All of the results for ambulatory encounters have been revised to exclude these few counseling encounters, and all related t-tests have been recalculated.

TABLE 6. The Percentage Distribution of Problems Reported During Ambulatory Non-Counseling Encounters Classified by Bodily Systems and Other ICD Categories* for Alcoholics and Controls

Problem Categories	Percentage Distribution of Outpatient Problems		
	Alcoholic Study Subjects	Controls	Relative Risks
Psychosocial problems	33.0(26.5)†	12.4(12.2)†	6.3(4.1)†
Cardiovascular problems	12.8	11.5	2.7
Respiratory problems	12.0	21.5	1.3
Symptoms and ill-defined conditions	11.8	7.0	4.0
Accidents, poisonings and violence	7.0	13.9	1.2
Musculoskeletal problems	7.0	14.1	1.2
Digestive problems	5.9	3.9	3.6
Endocrine problems‡	2.9	1.0	6.6
Skin problems	2.7	8.6	0.75
Urogenital problems	1.8	2.4	1.8
Sensory or nervous problems	1.1	1.6	1.7
Infectious diseases	1.1	1.7	1.6
Neoplasms	0.9	0.4	4.8
Congenital anomalies	—	—	—
Total	100.0	100.0	100.0
Weighted frequency scores§	28,928(22,873)†	12,200(12,180)†	2.4(1.9)†
Index of Dissimilarity: 31.1			

* The major disease categories in ICD-8 (The Eighth Revision of the International Classification of Diseases) are used.[24]

† Figures in parentheses are adjusted percentages, excluding the code for alcoholism (code #303) from among the psychosocial codes.

‡ Obesity, problem code 277, is classified as a psychosocial problem in this study.

§ Overall weighted frequency scores are obtained by summing the total occurrences of each problem code weighted by the number of cases with one or more occurrences of the code for all codes in each ICD category. Thus, for example, 24 alcoholics had one or more reports of the problem of hypertension during ambulatory encounters. The problem of hypertension was reported a total of 129 times for these 24 alcoholics. The weighted frequency score for the problem code, hypertension, is therefore 3096.

other utilization categories, both groups reported about the same ratio of problems to encounters (1.5 to 1 for alcoholics; 1.4 to 1 for controls).

Prevalence of Psychosocial Problems among Alcoholics

Table 6 shows the distribution of problems, classified by standard ICD categories,[24] that were associated with outpatient non-counseling encounters for the sample. Based on this table, the alcoholic study subjects were particularly prone to be seen by non-counseling providers in ambulatory encounters for "psychosocial" problems. These included anxiety, depression, family and marital problems, etc. Fully one third of problems reported by alcoholics for outpatient encounters were classified as "psychosocial," compared with 12 per cent of controls' problems. This difference is highly significant ($t = 5.39$; df $= 105$; $p = 0.0001$). Cardiovascular problems, respiratory problems and ill-defined conditions represent about 12 per cent each of alcoholics' outpatient problems. Study subjects were at nearly three times higher risk of cardiovascular problems and 30 per cent greater risk of respiratory problems. They were four times more likely to report ill-defined symptoms, including nervousness and fatigue, chest pains of unknown origin, etc. Alcoholics were about 20 per cent more likely to use outpatient services for accidents and related problems and for musculoskeletal problems than controls. The former were nearly four times more likely to report or be diagnosed with digestive problems and about seven times more likely to report endocrine problems. Sensory or nervous problems and urogenital problems were about 80 per cent more common for alcoholics, while neoplasms were five times more common among them. Finally, skin problems were 25 per cent more likely to be reported for controls. Overall, the alcoholics' relative risk of reporting outpatient problems in these categories is 2.4, and the index of dissimilarity is 31 per cent.

Three psychosocial problems are among the most frequently reported for alcoholic study subjects in the ambulatory utilization category; they are alcoholism, depression and anxiety. Nervousness/fatigue, although classified as an ill-defined condition, may be construed as an emotional problem. While this problem is among both alcoholics' and controls' most commonly reported outpatient problems, alcoholics were six times more likely to report it. Alcoholism, the most common psychosocial problem for alcoholics, was recorded at least once during outpatient encounters for 36 study subjects (about 40 per cent of the study group) and 4 controls (about 4 per cent of the control group).¶ Depression was reported 12 times more often among alcoholics, while anxiety was four times more frequent. Less difference was found for obesity, which was classified among psychosocial problems. Obesity, although twice as common among alcoholics, was the most common psychosocial problem for controls, affecting 19 of them. Other psychosocial problems showed even greater relative risks for alcoholics. Sleep problems and family problems were 15 times more likely among alcoholics. Marital problems were 12 times more frequent. Job and financial problems and other general psychiatric or emotional problems occurred 6 times more often among alcoholic study subjects. Parenthetically, alcoholics (seven) were found to be 100 times more likely to be diagnosed as having physical disorders of presumably psychogenic origin than controls (one). These findings support Boyajy's contention that alcoholics tend to utilize more services for illnesses with a "high emotional component"[19] than do controls.

The same trends are apparent for problems associated with inpatient hospitalizations. Table 7 shows that more than half of all alcoholics' problems in this utilization category were psychosocial, compared with 1 per cent of controls' problems. This difference is statistically significant at the 0.001 level ($t = 3.44$; $df = 105$). This result is consistent with the Genesee Valley study's finding of a very high risk of psychiatric hospital admissions for alcoholics compared with controls. According to Table 12, which shows the most frequent inpatient problems for alcoholics, alcoholism is the single most frequent such problem requiring hospitalization. Two other psychosocial problems, depression and suicidal ideation, are also represented among the most frequently reported codes for study subjects. Controls had *no* psychosocial problems among those most frequently reported or diagnosed.*

Among alcoholics' inpatient problems, accidents and related traumas constituted 20 per cent. They were four times more

---

¶ If HMO providers identify the problem of alcoholism among controls in the same proportions as alcoholics are identified by providers' entries in the medical records (40 per cent), then as many as 11 controls might be alcoholics. This does not invalidate the control group, since controls are not defined as nonalcoholics, but as HMO members who have not been identified by the counseling department as alcoholics during the pivotal one-year-and-three-month period.

* The most common psychosocial problem associated with alcoholics' hospitalizations was alcoholism, which was reported in conjunction with the inpatient encounters of 12 study subjects (WFS = 336). Depression was associated with 5 study subjects' hospitalizations (WFS = 40); suicidal ideation with 3 (WFS = 24); obesity with 4 (WFS = 16); and marital problems with 3 (WFS = 12). The only two psychosocial problems that were reported in conjunction with controls' inpatient encounters were obesity for one control and anxiety for one control.

TABLE 7. The Percentage Distribution of Problems Associated with Inpatient Service Utilization for Alcoholics and Controls, Classified by Bodily Systems Involved and by Other ICD Categories*

Problem Categories	Percentage Distribution of Inpatient Problems		
	Alcoholic Study Subjects	Controls	Relative Risks
Psychosocial problems	52.6 (23.2)†	1.3 (1.3)†	231 (63)†
Accidents, poisonings and violence	20.4	28.0	4.1
Urogenital problems	8.5	8.9	5.4
Digestive problems	7.2	8.9	4.5
Cardiovascular problems	4.2	26.1	0.9
Endocrine problems‡	2.3	1.9	6.7
Musculoskeletal problems	1.6	10.8	0.8
Neoplasms	0.9	—	—
Skin problems	0.8	0.7	7.0
Respiratory problems	0.6	1.9	1.7
Sensory/nervous problems	0.6	4.5	0.9
Symptoms and ill-defined conditions	0.3	3.2	0.6
Infectious diseases	—	3.2	—
Congenital anomalies	—	0.6	—
Total	100.0	100.0	
Weighted frequency scores§	879 (543)	157 (157)	5.6 (3.5)
Index of Dissimilarity: 52.7			

*See footnotes to Table 6 for explanations.

common than among controls. Urogenital problems and digestive problems represent 8 per cent and 7 per cent, respectively, of all causes of hospitalization for alcoholics. Alcoholics were at about five times higher risk of such hospitalizations than controls. Controls, on the other hand, were slightly more likely to receive inpatient services for musculoskeletal problems and for ill-defined conditions, while there is virtually no difference between the groups in reported cardiovascular and sensory or nervous problems. On the whole, alcoholics were nearly six times more apt to be hospitalized for all problems than controls. The index of dissimilarity for this distribution is 53 per cent.

The observed differences between study and control groups in the prevalence of psychosocial problems in both outpatient and inpatient service utilization are not attributable solely or even primarily to inflation of the figures by the overrepresentation of the problem code, alcoholism, among alcoholics. This is shown by the adjusted percentages and relative risks in parentheses in Tables 6 and 7. Strong group differences remain after exclusion of the "alcoholism" problem code. The prevalence of psychosocial problems among alcoholics is a real phenomenon, rather than an artifact of the coding.

Inappropriate Care-Seeking by Alcoholics

The finding that many outpatient noncounseling encounters by alcoholic study subjects involved psychosocial rather than strictly medical problems needs explaining. Since alcoholism figured prominently among alcoholics' psychosocial problems, some attention was clearly being paid to it. Primary-care providers at the HMO were directly addressing the alcoholism of about 40 per cent of study subjects, regardless of how it was made known to them. Knowledge may have come through the medical history, through contact with the spouse or other family members, through direct contact with an intoxicated subject, or through entries made in the medical records by counseling providers.

Based on their patterns of utilization, study subjects seem to have been burdening the medical-care delivery system by inappropriate care-seeking. Alcoholics tended to present in large numbers to non-counseling providers for problems that would have been more appositely handled by the counseling staff. Certainly, these alcoholics were not strangers to that department, having been selected on that basis.

What can happen when alcoholics are not treated for the primary problem of alcoholism is that providers may treat their symptoms and complaints in the same way that other anxious, depressed, nervous and fatigued patients may be treated, particularly by drug therapy. For patients' psychosocial problems, providers may prescribe certain central nervous system (CNS) depressants, including diazepam, oxycodone, chlordiazepoxide, and other drugs listed in Table 8. As they are alcohol-dependent, alcoholics may also become dependent on these prescription drugs. The *Physician's Desk Reference* (PDR) contains manufacturers' warnings that addiction-prone individuals, such as alcoholics, should be carefully followed in drug therapy with CNS depressants, because of the increased danger of drug overdose, due to the additive effects of these drugs with alcohol.[26]

In this sample, study subjects were much more likely than controls to be prescribed certain sedative, tranquilizing and other CNS depressant drugs during ambulatory non-counseling encounters. Table 8 shows the relative frequency with which alcoholics were prescribed a number of drugs that are potentiated by alcohol and carry manufacturers' warnings about the danger of overdosage and chemical dependency. In no case did controls receive more prescriptions for these drugs. Based on Table 8, alcoholics were about six times more likely to receive prescriptions for diazepam and chlordiazepoxide; they were about 12 times more likely to receive flurazepam and amitriptylene; and they were many times more likely to be given oxycodone and imipramine. Propoxyphene, which was three times more likely to be prescribed to alcoholics, is among the more dangerous of these selected drugs, if the PDR manufacturers' warnings are any indication.†

Diazepam was the most frequently prescribed drug, in this selected group of drugs, for both alcoholics and controls. Thirty-five alcoholics and 17 controls were prescribed the drug, according to medical record reports, during non-counseling encounters. This is not to say that the original prescriptions for certain drugs may not have come from the counseling department for some of their clients or that such drug prescriptions are not monitored and patients followed. The relatively high prevalence of drug overdoses and combined overdoses of alcohol and other drugs among hospitalized alcoholics, discussed below, should be viewed in light of these drug prescription patterns.

Related to the tendency of alcoholics to seek medical care inappropriately for emotional and psychosocial problems is the issue of the relationship of utilization behavior to length of time in counseling treatment. The question is whether study subjects who stayed in counseling treatment were the high utilizers of outpatient non-counseling services for psychosocial problems. This would be expected if high medical and high psychiatric utilization go together. On the other hand, the sample alcoholics who dropped out of counseling treatment might have been largely responsible for the prevalence of psychosocial problems reported or diagnosed by medical personnel. The subjects in this case would have been substituting primary medical care for counseling care. This and the related issue of whether HMO medical

---

† The *Physician's Desk Reference* (PDR) contains the warning, "Do not prescribe propoxyphene for patients who are suicidal or addiction-prone." Propoxyphene products ". . . in excessive doses, either alone or in combination with other C.N.S. depressants, including alcohol, are a major cause of drug-related deaths."[26]

TABLE 8. Numbers of Study and Control Subjects Receiving Prescriptions for Selected Sedatives, Anti-Depressants, Tranquilizers, and Other CNS Depressant Drugs During Ambulatory Non-Counseling Encounters and the Relative Risks Associated with Reports of Such Prescriptions

Selected Drugs	Numbers of Subjects Receiving Prescriptions				Relative Risk
	Alcoholics		Controls		
	Number of Cases*	WFS†	Number of Cases*	WFS†	
Diazepam	35	3255	17	544	6
Oxycodone	16	400	4	16	25
Flurazepam HCl	11	198	3	15	13
Imipramine HCl	11	187	0	0	—
Amitriptyline HCl	7	112	3	9	12
Chlordiazepoxide HCl	8	104	3	21	5
Propoxyphene	7	56	4	20	3

* Number of cases is the number of subjects for whom each drug was prescribed one or more times during outpatient encounters.

† WFS is the weighted frequency score, that is the number of subjects receiving each drug multiplied by the number of times each drug was reported as prescribed for all subjects combined.

services are used more appropriately and at reduced rates by alcoholics as a result of outpatient alcoholism treatment await further data analysis.

The Role of Accidents, Poisonings and Violence in Service Utilization

The predominance of psychosocial problems among alcoholics' inpatient problems masks other relationships. When psychosocial problem codes are excluded from alcoholics' inpatient diagnoses, the category of accidents, poisonings, and violence emerges as accounting for nearly half of all remaining inpatient problems. Urogenital and digestive problems also increase in importance in association with alcoholics' hospitalizations. These represent about one third of all inpatient problems, when the impact of psychosocial problems is eliminated.

That problems classified as accidents, poisonings and violence represent a larger share of both outpatient and inpatient problems for controls than for alcoholics requires some discussion. This is shown in Table 6. Accidents, poisonings and violence comprise 14 per cent of controls' outpatient problems, compared with 7 per cent of alcoholics' outpatient problems. Similarly, in Table 7, 28 per cent of all inpatient problems for controls were accidents, poisonings and violence, compared with 20 per cent of inpatient problems for alcoholics. This result is misleading, however, since comparison of the weighted frequency scores for the two groups shows that alcoholic study subjects were 20 per cent more likely to use outpatient services for accidents and related problems and were 300 per cent more likely to be hospitalized for such problems than were controls.

Table 9 shows the percentage distribution of particular categories of accidents, poisonings and violence, associated with inpatient utilization for alcoholics and controls. The two groups differ considerably in the problems that led most often to hospitalization. Fully 40 per cent of accidents, poisonings and violence associated with alcoholics' inpatient utilization were due to the adverse effects of medicinal agents. Nearly half were due to all types of adverse effects. By contrast, controls experienced *no* such problems. Lacerations and open wounds, especially of the head, neck and trunk, were frequent accidents associated with alcoholics' hospitalizations (16 per

TABLE 9. The Percentage Distribution of Problems Classified as Accidents, Poisonings and Violence Associated with Inpatient Utilization for Alcoholic Study Subjects and Controls

Types of Accidents, Poisonings and Violence	Percentage Distribution of Problems		
	Alcoholic Study Subjects	Controls	Relative Risk
Adverse effects of substances (medicinal and non-medicinal)	48.0	—	—
Lacerations and open wounds	15.7	56.8	1.1
Internal injuries*	10.6	2.3	19
Fractures	7.3	4.5	6.5
Superficial injuries	6.7	22.8	1.2
Contusions and crushing injuries	6.1	4.5	5.5
Sprains and strains of joints	5.0	2.3	9
Dislocations	0.6	4.5	0.5
Burns and effects of foreign bodies	—	2.3	—
Total	100.0	100.0	
Weighted frequency scores	179	44	4.1
Index of Dissimilarity: 63.4			

* This category includes intracranial injuries, internal injuries of the chest, abdomen, and pelvis, and injuries to the nerves and spinal cord.

cent), as were intracranial injuries, especially concussions (9 per cent), and fractures, especially of the skull, spine and trunk (7 per cent). Controls' most frequent accidents associated with inpatient utilization were lacerations and open wounds of the head, neck and trunk. These accounted for 57 per cent of all controls' accidents. Twenty-three per cent of their accidents were superficial injuries, compared with 7 per cent of alcoholics' accidents.

Clearly, the distribution of accidents for study subjects is powerfully skewed by the predominance of adverse substance effects. Within the adverse substance effects category, which accounts for such a high proportion of inpatient accidents and related problems among alcoholics, drug overdoses and combined alcohol and other drug overdoses are major contributors, constituting about 80 per cent.

When the relative risks for the two groups are examined, alcoholics were 6.5 times more likely to be hospitalized for fractures; 9 times more likely to be hospitalized for sprains and strains of joints; 19 times more likely to be hospitalized for internal, especially intracranial, injuries; and 5.5 times more likely to be hospitalized for contusions and crushing injuries, compared with controls. Both groups were about equally likely to be hospitalized for lacerations and superficial injuries. Controls, on the other hand, were slightly more likely to use hospital services for dislocations and for burns, although the numbers are very small.

The percentage distributions of particular accidents and related problems reported or diagnosed in conjunction with alcoholics' and controls' outpatient service utilization are given in Table 10. Of all accidents that occasioned outpatient encounters for study subjects, 59 per cent were sprains and strains of joints, 12 per cent were adverse effects of medicinal agents and other substances, 9 per cent were fractures, and 7 per cent were contusions and crushing injuries. Of controls' accidents associated with outpatient encounters, the vast majority (82 per cent) were sprains and strains of joints; 6 per cent were fractures and 4 per cent were adverse effects. There is more convergence between the distributions of accidents for which outpatient HMO services were utilized by alcoholics and controls than was the case for the inpatient comparison.

TABLE 10. The Percentage Distribution of Problems Classified as Accidents, Poisonings and Violence Associated with Ambulatory Encounters for Alcoholic Study Subjects and Controls

Types of Accidents, Poisonings and Violence	Percentage Distribution of Problems		
	Alcoholic Study Subjects	Controls	Relative Risk
Sprains and strains of joints	58.5	82.4	3.8
Adverse effects of substances (medicinal and non-medicinal)	12.2	3.9	3.8
Fractures	9.4	5.7	2.0
Contusions and crushing injuries	7.5	0.8	11.8
Lacerations and open wounds	6.0	3.8	1.9
Superficial injury	4.5	1.6	3.3
Dislocations	1.6	1.3	1.6
Internal injuries*	0.3	0.4	0.7
Burns and effects of foreign bodies	0	0.1	—
Total	100.0	100.0	
Weighted frequency scores	2040	1698	1.20
Index of Dissimilarity: 24			

* This category includes intracranial injuries, internal injuries of the chest, abdomen and pelvis, and injuries to the nerves and spinal cord.

This is shown by comparing the index of dissimilarity from Table 9—63 per cent—with that of Table 10—24 per cent.

Nevertheless, here again, in nearly every problem area, alcoholic study subjects utilized outpatient services for more accidents, poisonings and violence than did controls. The former sought outpatient care for twice as many fractures (2.5 times more skull, spine and back fractures); about twice as many lacerations and open wounds; 3 times more superficial injuries; 12 times more contusions; and 4 times more adverse effects of drugs and other substances. Controls, on the other hand, utilized outpatient services for 15 per cent more sprains and strains of joints and for slightly more internal injuries, specifically intracranial injuries, and more burns and effects of foreign bodies. Again, the numbers in these last categories are very small.

The fact that alcoholics were four times more likely than controls to be hospitalized for accidents, poisonings and violence and only about 20 per cent more likely to utilize outpatient services for such problems is puzzling. One possible explanation for this is that alcoholics may be going to hospital emergency rooms with problems that controls might take to the HMO. This is not likely, since the decision to seek hospital emergency room care is not usually made by individual HMO members, but is normally made by RIGHA providers, often by telephone. This process of permission-seeking is designed to reduce or eliminate inappropriate and uneven utilization of emergency room services. It insures that given similar problems, both study subjects and controls would be hospitalized at the same rate.

Another possible explanation is that the type of accidents and related problems that alcoholics have tend to be more serious and life-threatening than those experienced by controls. They are more likely therefore to require inpatient services. This notion is supported by the fact that "adverse effects," particularly drug and alcohol overdoses, were so frequently associated with inpatient encounters for alcoholic study subjects and *never* led to inpatient utilization by controls. The synergistic and potentiating effects of alcohol on other CNS depressant drugs are probably major causal factors here.[5,27] Such adverse effects were also associated with four times more outpatient utilization for alcoholics than for controls. This suggests an important way in which distinctly alcohol-related and alcohol-induced problems represent a burden on this health-care delivery system.

It is likely that alcoholics are reluctant to seek care for minor injuries for a number of reasons: because they do not want their alcoholism to be discovered in the process of explaining what happened; because being anesthetized and sedated by alcohol, they may hardly notice such injuries; and because being able to mobilize one's faculties and resources to seek medical care may be delayed by intoxication. By the time "morning comes," a laceration may be too old to suture, and the motivation to seek care may recede until a complication of the injury, such as infection, arises (Dr. LeClair Bissell, personal communication, August 14, 1980).

Two types of accidents do not appear to be directly or clearly alcohol-related. One is lacerations and open wounds, especially of the head, neck and trunk, for which alcoholics and controls were equally likely to be hospitalized. The other is sprains and strains of joints, for which controls were actually more likely to utilize outpatient services than were study subjects. Since many of the controls who suffered lacerations may have been involved in automobile and other accidents with alcoholics or under the influence of alcohol themselves, the role of alcohol here is difficult to determine. However, as noted, alcoholics may ignore many relatively minor accidental injuries for which outpatient care might be sought by others. Thus, the surprisingly high outpatient utilization by controls for sprains and strains may be attributable, in large part, to controls' tendencies to attend to such relatively minor injuries, before they are exacerbated or self-limited. Finally, it should be noted that alcoholics were hospitalized nine times more often than controls for sprains and strains, lending support to the idea that when alcoholics are injured and seek treatment for their injuries, their condition is more serious, and the services required are correspondingly more extensive and costly. In addition, if controls are generally healthier than alcoholics and if controls are less likely to utilize HMO services for psychosocial problems and for chronic and serious illnesses than are alcoholics, then it is understandable that accidents and related problems represent a higher proportion of problems associated with controls' service utilization.

There is debate over whether accidents, poisonings and violence are directly associated with alcohol, or whether, as some writers suggest, certain personality characteristics or some antecedent conditions give rise to both alcoholism and accident-proneness.[5] More research is needed in this area, but it seems likely that a combination of the anesthetic, sedative effects of alcohol, the possibility of potentiation of these effects, when alcohol is used with other CNS depressants, and the behavior associated with intoxication is central to explaining many types of accidental injuries, both unintentional and deliberate.[27]

Chronic Illness Among Alcoholics

A final brief insight may be gained into the issue of whether alcoholics utilize more HMO services, partly because they are more likely to be chronically or seriously ill than controls. Tables 11 and 12 provide lists of the most frequent problems, based on single, non-collapsed ICD codes, associated with outpatient and inpatient encounters, for comparison between study and control groups. They give some indication that chronic problems are likely to predominate among those most frequently reported for alcoholic study subjects. More acute problems seem to be listed among those most frequently recorded for controls.

Setting aside consideration of emotional problems, alcoholics tended to report more frequent chronic conditions, such as, diabetes, hypertension and obesity, than controls, judging by the magnitude of rela-

TABLE 11. The Most Frequent Problems Associated with Outpatient Non-Counseling Encounters for Alcoholics and Controls and the Corresponding Relative Risks

Alcoholics				Controls			
Problems in Order of Frequency	No. Cases*	WFS†	RR‡	Problems in Order of Frequency	No. Cases*	WFS†	RR‡
1. Alcoholism	36	3492	175	1. URI§	31	1395	—
2. Hypertension	24	3096	3	2. Sprains and strains of sacroiliac	19	1140	5.7
3. Nervousness/ fatigue	37	2331	6	3. Hypertension	12	1044	—
4. Obesity	27	1728	2	4. Obesity	19	855	—
5. Depression	31	1581	12.3	5. Allergic rhinitis	16	544	1.1
6. URI§	31	1519	1.1	6. Anxiety	16	400	—
7. Anxiety	25	1425	3.6	7. Nervousness/ fatigue	15	390	—
8. Diabetes	9	756	10.1	8. Bursitis or tendonitis	12	384	1.3
9. Sprains and strains of back and neck	15	615	4.5	9. Costochondritis, etc.	9	306	—
10. Costochondritis, osteochondritis, epicondylitis	13	598	2.0	10. Low back pain	10	290	0.74

* Number of individuals reporting one or more occurrences of the problem.

† Weighted frequency score.

‡ RR = Relative Risk.

§ URI = Upper respiratory tract infection.

TABLE 12. The Most Frequent Problems Associated with Inpatient Encounters or Hospitalizations of Alcoholics and Controls

Alcoholics				Controls			
	Problems in Order of Frequency	No. Cases	WFS*		Problems in Order of Frequency	No. Cases	WFS*
1.	Alcoholism	12	336	1.	Face or scalp lacerations	5	25
2.	Drug overdose	6	54	2.	Acute MI, heart attack	3	18
3.	Depression	5	40	3.	Angina pectoris	3	12
4.	Pregnancy	5	35	4.	Ruptured disc	2	12
5.	Face or scalp lacerations	5	25	5.	Pregnancy	3	9
6.	Suicidal ideation	3	24	6.	Gallbladder problems (cholecystitis, cholelithiasis)	4	8
7.	Diabetes	3	18	7.	Superficial injury of face, neck, or head	2	6
8.	Obesity	4	16	8.	Strep throat	2	4
	Fatty metamorphosis of the liver, hepatomegaly	4	16		Epilepsy, seizure disorder	2	4
	Female sterilization	4	16		Arterial embolism/thrombosis	2	4
	Concussion	4	16		Inguinal hernia	2	4
					Myalgia	2	4

* WFS is the Weighted Frequency Score, defined as the total number of occurrences of the problem code weighted by the number of cases or individuals whose records contained at least one mention of that problem.

ive risks for outpatient problems, shown in Table 11. Among the problems diagnosed most frequently for controls were such acute problems as upper respiratory tract infections (URI), sprains and strains of the sacroiliac region, and allergic rhinitis.

This conclusion is tenuous, partly because it is difficult to draw unambiguous boundaries between acute and chronic, and mild and serious conditions, and partly because obesity and hypertension were frequent problems for controls receiving outpatient care, as upper respiratory tract infection (URI) was a frequent problem for study subjects. These study subjects share many of the diagnoses listed as most frequent for alcoholics by Boyajy and by other authors.

Regarding hospitalizations, alcohol-related and/or chronic conditions seem to predominate among the most frequent problems for alcoholics (Table 12). Besides alcoholism itself, these include drug overdose and depression, and chronic conditions, such as, diabetes, fatty metamorphosis of the liver/hepatomegaly, and obesity. The controls' list again tends to include more acute problems among those most frequently associated with inpatient utilization. These conditions are strep throat, face and scalp lacerations, other injuries to the face, neck and head, ruptured disc, and acute myocardial infarction (MI). Anomalous conditions, such as angina and gallbladder problems, are also among controls' most frequent problems. Face and scalp lacerations and pregnancy, which was coded as a problem, are frequently associated with hospitalizations for both groups. If these single problem codes could be combined into meaningful categories to approximate degrees of severity or chronicity, a better test of this hypothesis would be possible.

Conclusions

Summary

The purpose of this study was to compare the utilization of HMO services by two groups of HMO members—a group of 106 identified alcoholics and an equal number of matched controls. Utilization and other data were collected from HMO medical records. Utilization was measured by numbers and proportions of encounters compared across four service utilization categories—laboratory and x-ray services, external referrals, hospitalizations and ambulatory non-counseling encounters. Differences between study and control groups in the distribution and frequency of problems or morbidities associated with each major utilization category were examined.

The results showed that, as measured by encounters, alcoholics utilized more of all HMO services studied than did controls. Study subjects were especially prone to utilize inpatient hospital services, even when hospitalizations for alcoholism detoxification and rehabilitation were excluded.

Some evidence was found for the notion that alcoholics utilize more HMO inpatient and outpatient services, because they are more likely to be physically and emotionally ill than controls.

In the first place, more problems were identified for alcoholic study subjects, compared with controls, in association with all types of encounters, in particular, with hospitalizations.

Secondly, significantly more of the outpatient non-counseling and the inpatient problems reported by the alcoholic study subjects were classified as psychosocial, compared with controls' problems. The importance of accidents, poisonings and violence increased as causes of hospital service utilization by alcoholics, when psychosocial problem codes were excluded. Alcoholics were found to utilize more outpatient and inpatient services than controls for such problems, although accidents represented a larger proportion of all problems for controls. Alcoholics were much more likely to be hospitalized for drug overdoses than controls. In general, study subjects seemed to seek treatment for injuries, when their condition was more serious and when the services required to treat them were more extensive and costly, than was the case for controls.

Finally, some tentative suggestions were made, based on a comparison of the most frequent outpatient and inpatient problems for alcoholics and controls. These support the notion that more serious or chronic problems tended to be associated with alcoholics' inpatient and ambulatory encounters. By contrast, more acute problems were noted for controls.

Further Research Needed

The second stage of this research project will address the issue of whether and to what extent the failure to engage alcoholic HMO members in counseling treatment is costly to the health care delivery system, viewed in terms of their propensity to use a disproportionate share of HMO services. How acceptance of alcoholism treatment affects service utilization patterns and the associated illness experience of alcoholics will also be studied. One question of interest, for instance, is whether study subjects' utilization of medical services for hypertension, diabetes, obesity and other alcohol-related problems, as well as for psychosocial problems, such as anxiety and depression, will decrease with treatment. Some of the policy implications of the results of offering third-party coverage for alcoholism treatment, particularly in HMOs will be discussed.

An investigation of the possible "dose-response-type" relationship of alcoholism and morbidity is also of interest. Using measures of the level and extent of alcohol involvement or impairment, including the duration of excessive alcohol consumption, Boyajy's contention may be tested. Do more severely involved alcoholics experience higher morbidity and use more services than other alcoholics?

A third area of vital concern is sex differences in alcoholics' morbidity and care-seeking. The general issue is whether the illness of alcoholism might act as a leveler of expected sex differences in service utilization. Male alcoholics may be found to utilize health care services at the typically high rates of females generally, excluding obstetric-gynecologic and other sex-specific problems and procedures.[28] If male and female alcoholics do utilize services at similar rates, then the question persists whether and how patterns of HMO use and associated morbidities may differ by sex. As mentioned earlier, how alcoholic women differ from their controls in utilization behavior and illness experience is an important aspect of this research issue.

The question of causality remains. How alcoholism affects the origin and/or the development of certain illnesses and injuries needs clarification. Some combination of factors seems to be operating here, including the physiologic effects of alcohol ingestion; possible alcohol-related conditions and habits, such as inadequate nutrition, exposure to the cold, lack of physical exercise, general neglect of health, and increased susceptibility or lowered resistance to infections[6]; combined alcohol and other drug use, including coffee or tea consumption, cigarette smoking, the use of tranquilizers, sedatives, and stimulants[11]; background or personality factors, including genetic predisposition; and social factors, such as anomie, social isolation, etc.[29] All of these variables may interact with alcohol and confound its effects on illness. Psychiatric problems, such as nervousness, anxiety, and depression, of which alcoholics have been shown to be at significantly higher risk than controls, may, according to Pell and D'Alonzo, ". . . lead to psychosomatic disorders and increased susceptibility to a wide range of diseases,"[6] especially those linked to stress. The issue is very complex, and relationships are difficult to disentangle. However, determination of the extent to which the excessive consumption of alcohol in and of itself is important in the natural history of certain illnesses, independent of other risk factors, is critical to our understanding of alcoholism and the other illnesses that may signal it.

Acknowledgments

The author gratefully acknowledges the expert technical assistance of Dr. Robert L. Stout, Assistant Professor (Research), Section of Psychiatry and Human Behavior, Brown University, Providence, Rhode Island and the generous advice and useful comments of Dr. LeClair Bissell, President and Chief Executive, Edgehill Newport, Newport, Rhode Island. Thanks are also due to Michael C. Gerhardt, the Occupational Alcoholism Project Director at Rhode Island Group Health Association, for his early conceptualization of the research problem and his editing suggestions, to the therapeutic staff of the counseling department and to the medical records staff for their cooperation, and to the staff of the Employee Assistance Program, especially to Wendy Woifson, for her most able and skillful research assistance, and to Sandra Peah, for typing the manuscript.

References

1. Berry R, et al. The economic costs of alcohol abuse and alcoholism—1977. Prepared for the National Institute on Alcohol Abuse and Alcoholism. Rockville: PB 278-606 AS.

2. JWK International Corporation. Benefit-cost analysis of alcoholism treatment centers. Volumes I and II. Prepared for the National Institute on Alcohol Abuse and Alcoholism (#ADM-281-75-0031) Annandale: 1976.

3. Deakin CML, Santamaria JN, Wilkinson P. The cost of alcoholism to the community. Med J Aust 1973;i:1305.

4. Wersinger R, Roberts J, Roghmann K, Wells S. The inpatient hospital and ambulatory utilization experience of an alcoholic population compared to a matched control group within an HMO. Presented at the twenty-eighth annual meeting of the Group of Health Institute, New York, June 18-21, 1978.

5. United States Department of Health, Education and Welfare. Alcohol-related illnesses. First Special Report to the U.S. Congress on Alcohol and Health. From the Secretary of Health, Education and Welfare. Rockville: U.S. Department of Health, Education and Welfare, 1971:45 (DHEW Publication #ADM 74-68).

6. Pell S, D'Alonzo CA. The prevalence of chronic disease among problem drinkers. Arch Envir Health 1968;16:679.

7. Barchha R, Stewart MA, Guze SB. The prevalence of alcoholism among general hospital ward patients. Am J Psychiatry 1968;125:681.

8. Wilkinson P, Kornaczewski A, Rankin JG, Santamaria JN. Physical disease in alcoholism—initial survey of 1,000 patients. Med J Aust 1971;i:1217.

9. Moore RA. The prevalence of alcoholism in a community general hospital. Am J Psychiatry 1971;128:638.

10. Brody JA, Mills GS. On considering alcohol as a risk factor in specific diseases. Am J Epidemiol 1978;107:462.

11. Williams AT, Burns FH, Morey S. Prevalence of alcoholism in a Sydney teaching hospital: some aspects. Med J Aust 1978;ii:608.

12. Penick EC, Norton JC, Sauers CA, Powell BJ, Read MR. Medical problems of the alcoholic in nonmedical treatment. Health Soc Work 1978;3:123.

13. DeLint J, Schmidt W. Alcoholism and mortality. In: Kissin B, Begleiter H, eds. Social aspects of alcoholism. New York: Plenum Press, 1976:275.

14. Pell S, D'Alonzo CA. A five-year mortality study of alcoholics. J Occup Med 1973;15:120.

15. Massé H. L'alcoolisme facteur de mortalité: son incidence sur les principales causes de décès. La Nouvelle Presse Méd 1972;1:1857.

16. Brock P, Coordinator. Summary. Report of the liaison task force on alcohol-related problems. Submitted to the President's Commission on Mental Health, February 15, 1978 (PCMH/P-78/26), 2078.

17. Putnam SL. Policy adoption—helps and hindrances: a study of factors associated with one measure of outcome success in an HMO-based occupational alcoholism project. Presented at the seventh annual meeting of the Association of Labor-Management Administrators and Consultants on Alcoholism, San Diego, October 3-6, 1978:217.

18. Boyajy TG. Alcoholism within prepaid group practice. Washington, D.C.: Group Health Association of America, 1978.

19. Boyajy TG. Demographic, patient functioning, and health care utilization profiles of alcoholics' pre-post treatment within three prepaid group practices. Presented at the twenty-eighth annual meeting of the Group Health Institute, New York, June 18-21, 1978.

20. Wan TT, Siofer SJ. Determinants of physician utilization: a causal analysis. J Health Soc Behav 1974;15:100.

21. McKinlay JB. Some approaches and problems in the study of the use of services—an overview. J Health Soc Behav 1972;13:115.

22. Anderson OW, Andersen RM. Patterns of use of health services. In: Levine S, Reeder LG, eds. Handbook of medical sociology. 2nd ed. Englewood Cliffs: Prentice-Hall, 1972;386.

23. Blue Shield of Rhode Island. Schedule of relative values for surgery; schedule of x-ray procedures; schedule of diagnostic laboratory services. Blue Shield of Rhode Island Manual, Providence, no date, 25.

24. Committee on Nomenclature and Statistics. Detailed list of major disease categories in ICD-8. DSM-II—Diagnostic and statistical manual of mental disorders. 2nd ed. Washington, DC: The American Psychiatric Association, 1968;83.

25. Blalock HM Jr. Social statistics. 2nd ed. New York, McGraw-Hill Book Company, 1972:223.

26. Baker CE Jr, Publisher. Physician's desk reference (PDR). 34th ed. Oradell, N.J.: Medical Economics Company, 1980.

27. Waller JA. Alcohol and unintentional injury. In: Kissin B, Begleiter H, eds. Social aspects of alcoholism. New York: Plenum Press, 1976;307.

28. Ashley MJ, Olin JS, le Riche WH, Kornaczewski A, Schmidt W, Rankin JG. Morbidity in alcoholics: evidence for accelerated development of physical disease in women. Arch Intern Med 1977;137:883.

29. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979;109:186.